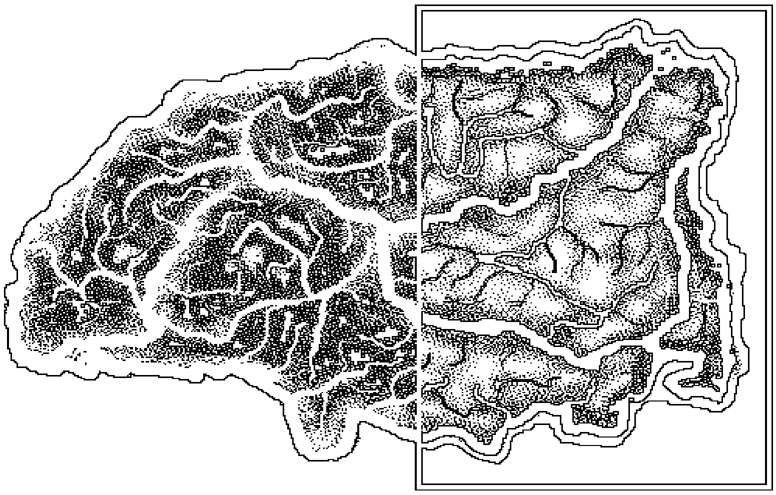
Complexity and Evolution

fundamental concepts of
a new scientific worldview



Francis Heylighen

Lecture notes 2014-2015

Preface

Over the last twenty years, there has been a silent revolution in science. Scientists in very divergent fields, from the natural sciences through information technology to the humanities, have steadily become more aware of the shortcomings of the accepted Newtonian paradigm. It has become clear to them that the underlying worldview is far too static, simplistic and reductionist to approach complex, dynamic phenomena, such as living organisms, people and society. The alternative approaches are silently converging under the header of "complexity science". This is a field that has not yet been clearly demarcated and can seem like a hodgepodge of metaphors, methods and models from a variety of disciplines and traditions. These different approaches do nevertheless share a distinct way of thinking that sheds an entirely different light on old philosophical and scientific problems. It is this way of thinking that I aim to present in a straightforward and coherent manner in this book, in the form of what I call the "evolutionary systemic worldview".

This book has developed out of my own research and aims to provide a synthesis of the different scientific (and philosophical) fields in which I work. I am a research professor at the Leo Apostel Centre of the Free University Brussels, where I lead a research group on "Evolution, complexity and cognition". The goal of the centre, as formulated by the well-known philosopher Leo Apostel, is the transdisciplinary integration of the varying disciplines of science and the humanities, and especially the formulation of integral worldviews. I am also the editor of the international Principa Cybernetica Project (http://pcp.vub.ac.be/), which aims to develop a worldview that is specifically based on the evolution of complex systems. The research has practical applications, such as the development of an intelligent, self-organising web. You can find more information about my work on my homepage, http://pcp.vub.ac.be/HEYL.html.

The content of this book reflects my own personal approach, not necessarily the scientific consensus. The reason for this is that the field of complexity and evolution is still largely at the research stage; the various ideas and approaches have not yet been systematically developed into a coherent theory. Several of the underlying theories (e.g. the biological theory of evolution) and especially several of the fundamental concepts (e.g. state space, feedback and fitness landscape) have however become solidly anchored in scientific thought. The book aims first of all to put forward these fundamental ideas in a clear and straightforward manner. Secondly, this book is a first attempt to integrate and generalise the approaches.

The book has been arranged as a textbook which has originally been used for the course "complexity and evolution" that I teach to first-year students in philosophy at the Free University Brussels. Consequently, the book requires no specialised prior knowledge other than that acquired in secondary education. Although the field of complexity science can frequently be quite technical, using involved mathematical formulas, the basic tenets are relatively simple. It is these basic tenets that I have aimed to explicate in a widely accessible manner, without trivialising them. This means that I explain the concepts as much as possible using concrete examples, illustrations and diagrams, and minimise the use of mathematical symbols. I consider it nevertheless a necessity to give an impression of how mathematics can help to represent complexity, using for example concepts such as state space and entropy. The mathematics required for this is however limited to elementary set theory as taught in secondary education.

To facilitate reading for readers without a scientific background, the text contains a number of sections (marked with *) and paragraphs (in smaller print) that discuss less important or more complicated applications, and that can be skipped without a problem. It is however worthwhile to read these sections, since they can contribute to the understanding of the basic concepts. Core concepts are marked **bold** in the text, so as to be easily recognisable. They can also be found in the alphabetical index at the end of this book.

Francis Heylighen

Table of contents

Chapter 1. Worldviews

1.1 Fundamental philosophical questions

For centuries, people have asked themselves fundamental questions about their existence and their place in the universe. These are essentially variations on the eternal "Why?":

- Why is the universe the way it is, rather than something else?
- Where does everything come from?
- Who are we?
- Where do we come from?
- Where are we going?
- What is the meaning of life?

Using a more modern, scientific terminology, we could summarise these questions as follows: how and why do complex, organized systems, such as living beings, humans and societies, arise? In which direction are they evolving? As we will see, these questions are also related to the classical questions that define the traditional domains of philosophy:

- What is? (ontology, metaphysics)
- What is true and false? (epistemology)
- What is good and what is evil? (ethics, axiology)

Together, the answers to all these questions determine a **worldview**, that is, a comprehensive philosophical system, a coherent vision of the whole (as defined by the well-known philosopher Leo Apostel and his collaborators in their book "Worldviews: from fragmentation to integration", see bibliography). A worldview gives meaning to our actions and offers guidance for understanding the world around us. A coherent worldview is especially important in the current age of ever faster scientific, cultural and social developments, in which all the old certainties are called into question. The confusion and fragmentation accompanying this often leads to anxiety and pessimism, as well as the need for the psychological support given by an easy to understand conceptual framework.

This framework is regrettably found all too often in fundamentalist ideologies or in irrational beliefs or superstitions. Science should be our pre-eminent weapon in the battle against irrationality and fundamentalism. Unfortunately, it seems that modern science contributes even more to the confusion with its deluge of often-contradictory observations and theories. One of the most important causes of this is the tendency of traditional science to divide all problems into sub-problems, which in turn are reduced to even more specialised sub-problems, and so on. The result is an abundance of very specialised information that appears to lack any semblance of coherence.

Even so, in modern science there are also contrary tendencies, which aim to reintegrate the different disciplines. The driving forces behind this are the concepts of (complex) *system* and of *evolution*. The idea of system focuses on the fact that elements do not stand on their own, but form a coherent whole. The idea of evolution notes that the different systems have a common origin and dynamics. These ideas are developed in more detail in this book, in the form of an **evolutionary-systemic** worldview able to tackle the fundamental questions. Indeed, all these questions can be reduced to questions about the origin and the further evolution of systems: atoms, molecules, galaxies, cells, plants, animals, humans, the brain, societies, civilisations, etc.

Recently developed scientific insights allow us to answer these questions. The answers are sometimes very abstract and general, sometimes more detailed. They can in principle be formulated in a mathematical form—although that is not always meaningful in practice. This scientific foundation means that they are more than just philosophical speculations. Indeed, they lead to concrete observations (of e.g. living beings or social developments) and applications (such as computer systems for complex problem solving, or the development and management of organisations).

The following sections briefly discuss the historical development of worldviews.

1.2 The religious worldviews

The first worldviews, from prehistoric civilisations until the Middle Ages, can be summarised under the denominator “religious”. Their answer to the question “why?” is simple: because God or the gods have wanted it. The gods are seen as personalised powers with their own goals and preferences, rather than as impersonal, natural mechanisms. The explanation for every natural phenomenon (such as lightning or the seasons) or social phenomenon (such as a famine or victory in war) is always divine intervention.

In relation to complex systems, this explanation has been formulated most clearly by the 18th century theologian, bishop William Paley, in his famous parable of the watchmaker. Imagine that while walking you find a complex, organised system, such as a watch. Where does it come from? It cannot have accidentally come into being, like a rock or a small heap of sand. Since it has clearly been assembled in an intelligent manner, it must have been developed by an intelligent being, in this case by a *watchmaker*. The different living creatures around us are such systems. They must therefore have been designed by an intelligent being: *God*.

The problem with this explanation however is that it is not a real explanation. After all, we still do not know why God has created things in exactly that way. Neither do we know who or what created God himself. Finally, this explanation has no predictive powers, since it is by definition impossible to understand the intentions of God or to find out what his plan is.

1.3 The Newtonian worldview

During the Renaissance and the following period of the Enlightenment, purely religious explanations began to be replaced by scientific explanations. This culminated in the 19th century in a new comprehensive worldview, which we will call **Newtonian** or mechanical, because it is based on the theory of classical mechanics, of which the founding principles were formulated by the physicist Isaac Newton.

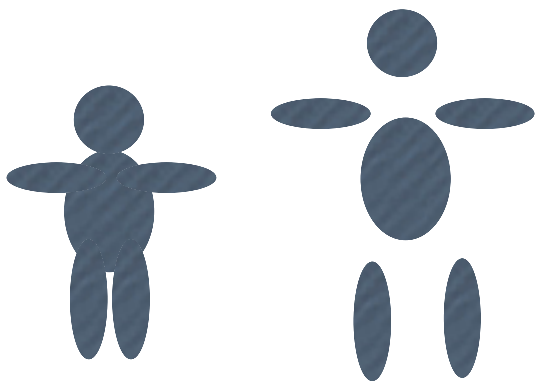
The Newtonian view can best be summarised as the **Clockwork Universe**: the cosmos runs like a mechanical clockwork. All the cogs fit perfectly together. The clockwork mechanism runs perfectly regularly and predictably and never shuts down. The inspiration for this view comes from the movement of the planets around the sun, which is indeed very regular and predictable. God may have created this "clockwork" once and set it in motion, but he is no longer concerned with it. Divine intervention is therefore no longer necessary to explain specific phenomena.

This new insight can be illustrated with a famous anecdote about the mathematician Laplace—after Newton one of the most important founders of the mechanical worldview. When Laplace set out his theory about the operation of the universe to Napoleon Bonaparte, the latter asked him what God's role was in this construction. Laplace answered simply: "I have no need for that hypothesis."

On the other hand, the importance of this insight can be illustrated by the different roads followed by Western and Islamic science. Although in the Middle Ages, Islamic science was way ahead of its European counterpart, its development petered out around the 11th—12th century. A plausible explanation for this is the theological doctrine of "occasionalism", which became dominant in the Arab world around that time. According to this philosophy, God intervenes at every moment in every process, and a cause produces an effect only because God has wanted it so. If one takes this philosophy literally, it no longer makes sense to search for scientific laws: after all, God does what he wants and thus makes every human prediction unreliable...

The origin of Newtonian science, on the other hand, can be found in freemasonry, which was involved in the creation of the very first scientific society, the Royal Society of London, founded in 1660, of which Newton was a member and later (1703-1727) chairman. The freemasons saw God as the Great Architect of the universe. This implies that God has designed the world according to rational, geometric principles. Once the world was created, he no longer had to intervene. The construction plan that God used is in principle understandable with the help of scientific observations and reasoning. It can be formulated in a mathematical form, as Newton did in his most famous work *Philosophiae Naturalis Principia Mathematica* ("Mathematical Principles of Natural Philosophy").

The influence of this mechanical worldview on modern society, especially during and after the Industrial Revolution, was so great that many still see it as *the* scientific worldview. Yet the underlying theory has become outdated through 20th-century developments, such as quantum mechanics, the theory of relativity and chaos theory (see further). Even so, it offers a simple, logical and coherent vision that remains useful in certain well-defined fields. Here are the basic principles of this worldview:



- Reductionism

All phenomena or systems (planets, living beings, societies, ...) can be analysed or reduced to their individual constituents: atoms, or, in more recent approaches, elementary *particles*. These are permanent, indivisible pieces of matter (*materialism*). They move through space, propelled by forces. All changes that we perceive around us can be reduced to the effect of such movements, which do not make any substantial alterations, but merely shift the positions of the components of the system.

- Determinism

All movement, and thus all change, is subject to *natural laws*. These laws are absolute and permanent. If you know the forces, velocities and positions of the particles at a certain time (which together define the **state** of the system), you can *completely predict* their movement by applying these laws. There is therefore no uncertainty concerning what will happen in the future: everything has already been decided (determined). There is no freedom to change or influence the course of events.

If we describe the state s as a function of time t, then the system follows a trajectory in the space of states (see chapter 8). In order to determine the state at all later points of time ($s(t > 0)$), it suffices to know the state in the present ($t = 0$).

- Reversibility

Trajectories cannot only be extended toward the future ($t > 0$), but also toward the past ($t < 0$). The state of a system can thus be determined for any point of time in the past. Every movement is reversible. In a sense, there is no real difference between future, present and past. Nor is there any *progress* in evolution, because change has no preferred direction: every movement could just as well have happened in reverse. A planet can for example orbit the sun in clockwise or anticlockwise direction; a canon ball can be shot from left to right or from right to left: according to the Newtonian laws of nature, both possibilities are equivalent.

- Weaknesses

Although it is more profound and more practical than the religious worldview, the Newtonian worldview still does not offer a complete explanation: the fundamental elements (space, time, particles, forces, laws) are after all postulated *a priori*, without real motivation. This worldview does not have room for purpose, value, or meaning: everything is reduced to purposeless mechanisms. Nor is there space for creativity, novelty or surprise.

In practice, this reductionist approach only works for simple mechanical phenomena: the trajectory of cannon balls, billiard balls, or planets around the sun, ... More complex systems, such as living beings, humans, or societies, are not explained. Quantum mechanics has shown that it fails even for the simplest phenomena, such as atoms or particles. These are after all subject to Heisenberg's uncertainty principle, which states that the movements of microscopic particles are intrinsically unpredictable. To conclude, the Newtonian worldview is completely unsuitable to explain life, people, or society, even though it helps us to understand some of their physical aspects (such as the effect of gravity on our body).

1.4 The evolutionary systemic worldview

Since the development of the theory of evolution in the mid-19th century, and especially since the development of cybernetics, systems theory and self-organisation in the mid-20th century, the first signs of a new worldview have appeared, removing the shortcomings in the Newtonian worldview. We will briefly list the innovations of this evolutionary systemic worldview (ESW) and analyse and motivate them in depth in the rest of this book.

In the ESW, there is no longer a need for God as the creator of the Universe. Complex organisation arises *spontaneously*, through accidental combinations of elements, without the need to appeal to an intelligent designer or "watchmaker". It is natural selection that counts: which combinations will continue to exist and which will not? This is generally not pre-determined, but depends on accidental, unpredictable factors. People are therefore free to make their own choices: after all, there is no determinism or predestination. Neither is there guidance, reward or punishment from above (God). These choices have to be viable, though, because otherwise humankind will be eliminated.

The theory of evolution allows us to explain the origin of all complex phenomena by evolution from earlier, generally simpler, systems. Bacterial cells have for example evolved from chemical cycles, complex cells evolved from bacteria, multicellular organisms from complex cells, complex animals from simpler animals, humans from animals, societies from groups of people. Post-modern philosophers have announced the "end of the Great Narratives", meaning that all explanatory models, scientific or religious, are relative. According to the ESW there certainly is a universal narrative or an origin story, but it is far more comprehensive and complicated than the traditional myths.

The ESW is intrinsically coherent. There are no strictly separated phenomena or categories, such as matter and mind, or space and time: every phenomenon is after all connected to, and arises from other phenomena. These interactions make the whole into more than the sum of its parts. There is an essential continuity between human, animal, plant and mineral.

The ESW vision is fundamentally optimistic. Its essence is self-organisation, or the spontaneous development of more complex and better adapted systems. Errors or negative developments do not persist, because they are eliminated by natural selection. "Nature" is creative and will sooner or later find a solution. Even if you leave things to their own devices, everything will eventually work out (although it is of course better to intervene sooner rather than later). In societies too, there is a clear trend towards progress. There is however no ultimate goal or endpoint to this evolution: everything can always be improved.

Although evolution is unpredictable, it does have a preferred direction: increased fitness (see further). This can help us to understand and direct our future, and to avoid choices that are not viable. Fitness is not a goal, but an implicit value of all life and matter. In a sense, this gives "meaning" to life.

We will now show the historical development of the evolutionary systemic worldview through the introduction of the fundamental ideas of each of the complementary developments that form the foundation of the ESW: the theory of evolution, self-organisation and chaos theory, systems theory, cybernetics, and complex adaptive systems.

Chapter 2. The theory of evolution

2.1 Darwin and the origin of species

In the religious worldview (and implicitly also in the Newtonian one), it was assumed that the species into which plants and animals are divided are unchangeable: the ancestors and descendants of cats are always cats—not dogs or cows—and their predecessors must therefore also have been cats. Cats can thus not have evolved from another species of animal, such as tigers. God has created the different species as they are now, and this is how they will remain.

This assumption was challenged when palaeontologists found fossils of no longer existing animals in old geological layers. It turned out that the older (deeper) the layer was, the more "primitive" (simpler) the fossils were. This impelled biologists to consider the mechanisms that could be responsible for this development from simple to complex species. Of the different explanations, it was eventually the theory of evolution by natural selection, discovered by Charles Darwin, that proved the most plausible.

This theory was initially very controversial, since it contradicted the religious worldview, as well as the Bible, which states that God created humans and animals in a single day. Darwin's fear of this controversy caused him to wait decades before he published his theory, until his colleague Wallace came up with a similar idea with which he would have taken all the credit. Even after publication Darwin kept himself aloof from the discussion as much as possible and left it up to others, such as Huxley, to defend his theory. Even though we may find it hard to imagine such an emotionally charged discussion, the theory of evolution remains anathema to certain groups of fundamentalist Christians and Muslims, the so-called "creationists", who take the creation story literally. This is particularly problematic in the USA, where the statistics show that more than half of the population believes in some form of creationism and only a third believes in the theory of evolution. (On the other hand, the Catholic Church has now officially accepted the theory of evolution as being scientifically supported.)

These creationists are strongly driven (and sometimes very creative) in their formulation of arguments to refute the theory of evolution—although upon closer inspection these arguments prove simplistic and misleading. The presence of simpler organisms in deeper layers is for example explained through the Biblical flood story: when the waters rose, the primitive animals would have been the first to drown, while more sophisticated animals would have been smart enough to find safety on higher lands, which only flooded later. It is self-evident that evolutionary biologists consider such an explanation ridiculous.

The theory of evolution also continues to encounter incomprehension from many people who may or may not be religiously or philosophically inspired and who cannot accept that blind, impersonal mechanisms can bring about functional order. Some of them have tried to prove scientifically that the complex systems we see around us would not have been able to evolve according to Darwinian mechanisms, because the likelihood that such a process would end well is far too small. This approach has recently led to a more sophisticated variety of creationism, the so-called "intelligent design" theory, which states that an unspecified intelligent power has been intervening to ensure that evolution would run smoothly. We will analyse this argument in detail at a later point.

Many progressive thinkers and social scientists are also critical about Darwinian explanations. This is however mostly based on a misunderstanding, where the theory of evolution is held responsible for the outdated ideology of "social Darwinism". This ideology applied Darwin's idea of "survival of the fittest" to society, to find a justification for a class society in which the poor or handicapped are left to their own devices. More recent applications of the theory of evolution however find that social systems should evolve towards more mutual help and solidarity in order to be successful, as we will demonstrate in some detail.

The theory of evolution is generally accepted within the natural sciences, although some scientists note that we should take other mechanisms into account besides natural selection, such as self-organisation and symbiosis. We will also discuss this criticism in more detail in later chapters.

2.2 How does evolution work?

Darwin's inspiration for his theory of natural selection came from the way horticulturists or animal breeders cultivate new breeds or varieties of plants and animals. There are for example hundreds of dog breeds, from the tiny Chihuahua to the gigantic Irish wolfhound, and from the hulking, compact bulldog to the slender, elegant greyhound. These are all descendants from the same wolf-like ancestor. If a breeder wants to create a breed with a special property (e.g. not barking), he starts with an existing group of dogs, of which he selects those that most closely approximate that property (i.e. those that are least likely to bark). Among the descendants of this first generation, some will bark more, others less. The breeder again selects those that bark least and makes these produce a new generation of descendants. From these, he again selects those that bark least, and so on. After many generations, he will eventually have a group of animals that never bark. The further descendants of this group will generally inherit this tendency not to bark from their parents. Thus a new, non-barking dog breed has been created.

Darwin's ingenious idea was the generalisation of this mechanism of artificial selection for a situation in which there is no breeder to select certain properties. Darwin noticed that there is selection in nature as well: generally, an organism has many more descendants than can survive. There is after all always a limited supply of food, and there are many diseases, predators and other dangers. Thus, most organisms die before they have had a chance to reproduce. Those who do manage have been selected from the initial group; the others have been eliminated. This selection happens spontaneously, by nature. That is why this mechanism is called **natural selection**. Nature does however not select for a specific property, such as non-barking, but for the general capacity to survive and reproduce—what we will later call **fitness**. This capacity will depend on the specific environment, and will require different properties in different environments.

Let me illustrate this with the classical problem of why giraffes have such a long neck. We start out with an ancestor of the giraffe, an antelope-like animal that still had a short neck. The descendants of this proto-giraffe all have slightly different genes, each a specific combination of the genes that they inherited from the father and mother. Some of these genes give rise to a slightly longer neck than others do. Giraffes with a longer neck can eat leaves from taller trees and thus have more food at their disposal. In times of famine, when nearly all the trees have been stripped of their leaves, only those descendants with the longest neck get enough food and so survive. Their descendants inherit the genes for a longer neck, but again with slight variations between the different offspring of a given set of parents. Of these, those with the genes for the longest neck will again have the biggest chance to survive. Thus in consecutive generations ever-longer necks are selected for.

Now suppose that some of these proto-giraffes have spread to an environment where there are no trees, but only low bushes. In this environment a long neck has no advantage, quite the contrary. Natural selection here will rather lead to the capacity to eat close to the ground. This group of proto-giraffes will therefore evolve to a variety with a short neck and legs and gradually come to differ more and more from their cousins with a long neck. After several hundreds of generations, the difference will be so large that we distinguish the two varieties as two species, which because of physical differences can no longer mate with each other and thus can no longer interbreed or produce intermediary forms. In this way, different species arise, each adapted to their own environment.

From these examples, we can deduce the general mechanism:

- blindly try out many variations on a certain base form
- eliminate the variations that do not work as well or that have not adapted as well (mistakes)
- keep and/or reproduce those that work better
- start again

This continuing trial-and-error procedure ("algorithm") spontaneously leads to constant, irreversible improvement, during which the system adapts ever better to the requirements of the environment. There are two essential components to this mechanism:

- **Variation**: the individuals upon which selection acts have to be different, and every new generation has to give rise to new differences. In biological evolution, these variations are generated through mutations (accidental errors in the replication process) in the genes, and through recombination of the genes of the father and mother.
- **Selection**: the "best" variations have to be selected. If something works, it is retained; if it does not, it is eliminated. The extent to which something works, or is adapted, is called **fitness**.

2.3 The increasing importance of the evolutionary approach

The idea of evolution through variation and selection is universal and is not restricted to biology or the origin of species. An increasing number of scientific disciplines appreciate its importance and use. One could even think that Darwinism has become a fashion over the last couple of years, if it were not for the fact that this trend has very deep roots. I will discuss several examples below:

- Psychology

It turns out that our thought processes and emotions can be explained by assuming that they are the result of selection for fitness. The emotion of jealousy, for example, causes men to keep an eye on their partner, to ensure that the children of their wives are also their own children and not those of a rival. Men who did not know this emotion had fewer descendants and their genes were thus selected against. The attraction of men for women with long legs and noticeable breasts can be explained because these are signs of sexual maturity and thus of the ability to produce descendants. Pre-pubescent girls on the other hand have short legs and flat chests, which is experienced as less attractive.

When it comes to thought processes, it turns out that people often make mistakes in logical deductions (syllogisms), which is understandable since natural selection has not prepared our brains for this kind of abstract reasoning. However, if we formulate the syllogism as a social problem, everyone immediately gives the right answer. Social interactions are after all very important for survival in groups, so our brains are especially adapted to deal with these kinds of problems.

- Sociology

"Sociobiology" explains the development of social systems by looking at the way these systems add to the fitness of their members. Vampire bats that have managed to suck a lot of blood, for example, donate a portion of their yield to their less fortunate colleagues with the expectation that they in turn can later count on *mutual aid* or solidarity. Thus these bats have a lower risk of starvation in times of scarcity. Ant colonies are one massive, integrated society in which every worker is willing to give her life for the colony. The reason is that all of the colony's members are children of the same "queen" and therefore largely share the same genes. These genes have been selected to maintain the entire family rather than a single individual. We will discuss these applications to social systems in more detail below.

- Medicine

Our susceptibility to and immunity against illnesses can be explained through natural selection of both humans (who need an efficient defence mechanism to survive) and pathogens (which need to be as contagious as possible in order to reproduce widely, but need to keep their host alive as long as necessary to enable the infection of others). The AIDS and common cold viruses are for example so successful because their victims do not become bedridden immediately, but are given the chance to spread the virus via contact with others (through sex or through sneezing near others). On the other hand, it is not very likely that the Ebola virus will spread widely, because infected people become deadly ill almost immediately.

Another example: dioxins are incredibly poisonous for laboratory animals, but much less so for humans. After the Seveso catastrophe of 1976, when an explosion in a factory in Italy released massive amounts of dioxins, no human being became deadly ill (although a few had developed skin problems). Animals such as rabbits however died in great numbers. A plausible explanation is that humans have been using fire for hundreds of thousands of years, and wood smoke does contain many dioxins. Genes that were not resistant to this have been selected out by now.

- Economics

Competition between businesses, where the most successful grow and are imitated and the less successful disappear, shows great similarity to natural selection between organisms. Here, the market plays the part of the environment that decides which businesses are adapted to the game of supply and demand and which are not. New businesses with new ideas or products are constantly being founded, while existing businesses regularly test new methods, services or technologies (variation). Most new businesses will eventually go bankrupt (natural selection). The most successful entrepreneurs are often those who find an as yet unexploited **niche**, that is, a way to sell products or services for which there is a large demand, but where there are as yet no competitors.

- Culture

Cultural phenomena such as fashion, traditions, scientific theories and religions can be described as ideas that are transmitted from person to person. Such pieces of information that “reproduce” themselves are in that way analogous to our genes and are therefore called *memes*. Memes are engaged in constant competition: we are continuously exposed to much more information than we can remember or pass on to others. We will remember and pass on only a small percentage of the ideas that we hear, read or see. Consequently, there is a strong natural selection of memes. Additionally, we generally introduce small variations when we for example tell an anecdote, a joke or pass on a rumour to someone else. This combination of variation and selection causes memes to adapt to society’s preferences.

The “fittest” memes are spread widely and eventually become myths, traditions or generally accepted knowledge—even if they are not based on fact. An example is the so-called Mozart effect: the recurring idea that your baby becomes more intelligent if it regularly hears classical music. This urban legend comes from the mistaken representation of an experiment in which *adults* listened to music and directly afterwards appeared to score better on certain psychological tests (perhaps simply because the music had relaxed them). It is only a small variation from adults to children, and from children to babies. Moreover, since parents like the idea that a simple intervention such as putting on a Mozart CD could make their baby smarter, this latest version has become very popular in newspapers and magazines. One journalist cites the other without anyone wondering about the scientific basis of this allegation.

- Computing

Computers can solve complex problems through the random generation of possible solutions, retaining only the best ones and allowing these to reproduce. During this process, variation, analogous to that in biological evolution, is introduced. Every potential solution is presented as a string of letters or bits, similar to the DNA strings that store an organism’s genetic information. During “reproduction”, several letters are randomly replaced by others (mutations). On the other hand, strings of two solutions (a “parent pair”) are recombined into a new string, which can for example consist of the first half of the first string followed by the second half of the second string. This is equivalent to sexual reproduction. The cycle is repeated so that ever more generations arise, of which only the best are allowed to reproduce. This continues until a satisfactory solution has been found. The method is called *genetic algorithms* or *evolutionary computation*. It turns out to be very useful to solve all sorts of practical problems that are too complicated for classical methods, such as the development of a “nervous system” for an autonomous robot, or an aerodynamic shape for a car.

- Chemistry

Through generating a variety of candidate molecules and repeatedly filtering out the best, it is possible to create molecules with specific properties. This is directly applicable in pharmaceutics: a medicine is a molecule that needs to suppress or stimulate a certain biological function (e.g. pain or fever). This is typically done through filling a “receptor” for certain signals in our body and thus blocking or facilitating the stimulation of that receptor by for example hormones. Receptors typically have very complex shapes, which makes it difficult to find a molecule that will fit exactly like a key in that lock. If we let candidate molecules pass receptors in great numbers, we will automatically find those that are adapted (those that stick to the receptor) and those that are not (those that are flushed away). Eventually, the best ones can be manufactured in great numbers to be sold in pill or drop form as a new medicine.

2.4 What is lacking in the evolutionary approach?

Despite its many successes, the Darwinian approach on its own is still too restricted to really generate a new worldview.

On the one hand, the theory is too reductionist. Organisms (in Darwin's original theory), genes (in the more modern, "neo-Darwinian" theory) or memes (the cultural variant) are seen as the primitive elements or "units of selection", to which everything should be reduced. Consequently, there generally is little attention for the interaction between these units (e.g. symbiosis of different organisms, or cooperation between genes), or for the systems (such as societies) that they form together. Instead, the complex influence of a multiplicity of other units is reduced simply to the influence of the "environment".

On the other hand, Darwinism does not really explain how complexity arises. Adaptation or fitness can after all be reached by both very simple systems (viruses, bacteria) and very complex ones (humans, societies). Increased fitness does not therefore imply increased complexity, and many biologists, among whom the palaeontologist Stephen Jay Gould, have argued that increased complexity such as that we see in fossils is only an accidental side effect of variation and selection, not a fundamental mechanism. My thesis in this book is exactly the opposite: evolution automatically produces more complexity, and those cases in which this does not appear to happen (e.g. viruses) are the exception rather than the rule.

To tackle these problems fundamentally, we need to add several new concepts to that of the theory of evolution, such as self-organisation, emergence, and co-evolution. To develop a good understanding of these, we will now discuss some complementary approaches.

Chapter 3. Self-organisation and chaos

3.1 Introduction

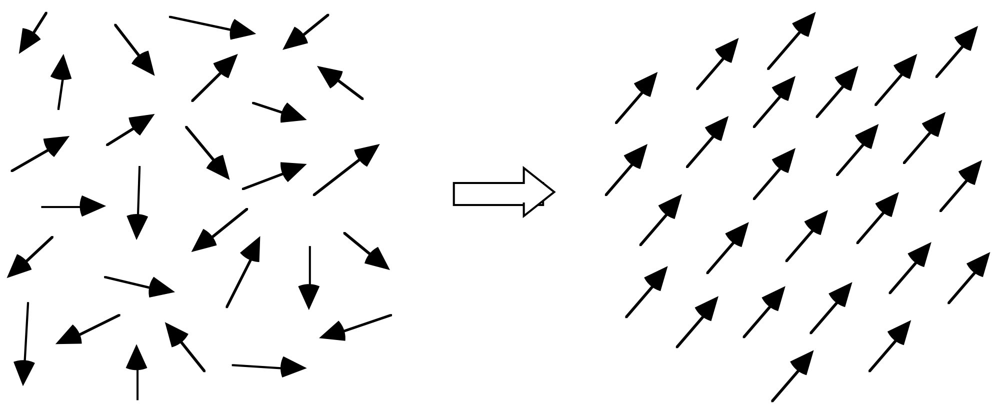
Nature provides plenty of examples of the spontaneous emergence of ordered structures, such as salt crystals in a solution, snow crystals in the atmosphere, or ice flowers on a window. No intelligent designer or "watchmaker" is involved, but neither is there clear influence of the environment, such as in Darwin's theory of evolution. Therefore, this phenomenon is called **self-organisation**.

The second law of thermodynamics (see 10.3) makes this even more paradoxical, because this law states that the entropy ("disorder") of an isolated system can only increase, not decrease. This corresponds to our daily experience: if we do not intervene, disorder increases: a room tends to dirty and messy, machines wear out or break and do not repair themselves.

This paradox in thermodynamics has been researched most thoroughly by Ilya Prigogine of the Free University of Brussels, who in 1977 got the Nobel Prize in Chemistry for his work. The most important contribution of Prigogine and his collaborators—the Brussels school of thermodynamics—is the concept of **dissipative structures**. These are forms of spontaneous order or structure, which perpetuate themselves by exporting ("dissipating") entropy. Using several examples, I will illustrate Prigogine's and others' insights on self-organisation, and the related phenomenon of chaos.

3.2 Examples: magnetization and Rayleigh-Bénard convection

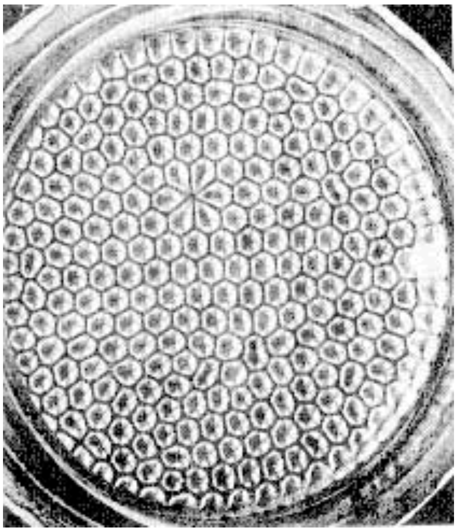
- Magnetization



Probably the simplest example of self-organisation is the magnetization of materials such as iron. You can produce magnetization by sliding a magnet over a pin, after which the pin itself becomes magnetic and able to attract other pins. The explanation is straightforward: magnetisable matter consists of molecules that each have an individual magnetic field, with a specific direction (shown with an arrow in the illustration below). These molecules can be seen as miniscule magnets. Initially, all these magnets point in different, random directions (as shown left on the illustration). This way, the varying magnetic fields or attractive forces negate each other, so that the total magnetism is zero. After magnetising the material, for example by subjecting it to an external magnetic field, all the little magnets point in the same direction (as shown right on the illustration). Thus the magnetic fields reinforce each other, producing a clearly observable total attractive force.

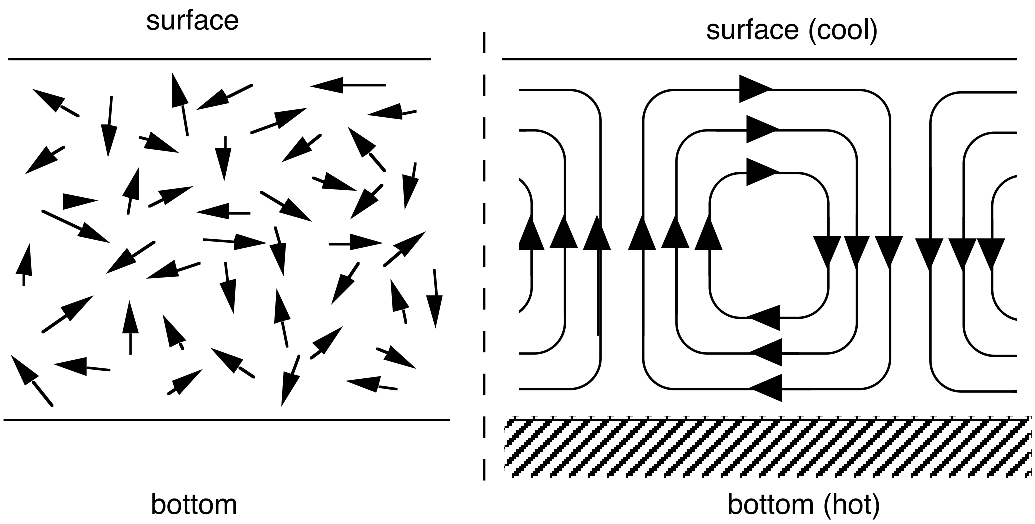
The interesting thing is that in certain circumstances magnetization can also arise spontaneously, without the interference of an external field. This is a clear example of self-organisation: the initially unordered little magnets (left) align themselves (right). The reason that the magnets prefer this ordered configuration is that magnets that point in opposite directiosn (e.g. the north pole of one magnet against the north pole of another) repel each other. There is no repulsion when they are all pointing in the same direction. Then the system is balanced.

- Rayleigh-Bénard convection



A somewhat more complex phenomenon is the development of so-called Bénard cells or rolls in a fluid. The fluid is heated evenly from below (for example on an electric stove), while it cools evenly from above (for example by contact with the air). (You can try to reproduce the experiment with a flat pan in which you pour enough oil to cover the bottom.) If the difference in temperature between bottom and top is large enough, you will see the development of a honeycomb pattern of hexagonal “cells” (see photo), or a striped pattern of parallel “rolls”. Below I will discuss the case of the rolls, because that is the simplest.

Just like the magnetic material, the fluid consists of molecules. However, since this is not solid matter, these molecules are in constant movement. Usually this movement is random, and every molecule moves in its own direction, independent of the others (shown by the arrows in the illustration, left). Therefore these molecules constantly collide with each other, with the result that the net movement is zero: the fluid is motionless.



But when the bottom is heated, the fluid here starts to expand and thus becomes lighter. The fluid at the top is colder and therefore heavier. The warm fluid will normally rise and the cold fluid will sink. This causes a problem, since the warm fluid is trying to work itself upwards, while in the same area, the cold fluid wants to go down. This will only happen if some form of coordination appears, so that these movements do not hinder each other. What happens is that in one place all molecules will make an upward movement until they reach the top. There, they cool down, after which they will they will sink again a little further. At the bottom, they heat up again, move back to their old place, and rise, after which the cycle repeats itself (right in the illustration). The net result is a sort of rotating current or "roll" of synchronised moving molecules. The current caused by the differences in temperature is called "convection". The entire fluid starts splitting into a series of parallel rolls. A roll that is rotating clockwise is always followed by a roll that is rotating anticlockwise (see illustration). Again, we recognise self-organisation: uncoordinated molecules spontaneously order their movement until they are all synchronized.

3.3 Global order

If we want to understand in more detail how such systems come to self-organise, we have to start with the interactions between the different components (the molecules in our example). Every molecule interacts (through its magnetic field or through collision) initially only with molecules in its immediate neighbourhood, and is independent of those molecules that are further away. Through these interactions or other accidental movements it can however happen that neighbouring molecules become aligned, i.e. point or move in the same direction. The other molecules in that neighbourhood will then tend to align or adapt themselves to those molecules, to avoid collision or repulsion. These newly aligned molecules will in turn align the molecules in their environments. In this way, the order or alignment propagates from molecule to molecule until it eventually covers the entire system. Thus, an accidentally arising, *local* order spreads through the material, resulting eventually in a *global* ordering of all molecules.

This order now directs or controls the behaviour of all components: individual molecules can no longer afford to go against the order (e.g. moving downwards if all the others nearby are moving upwards, or pointing left if the entire magnetic field is pointing right), because the collective influence of the other molecules is too strong. This order or control is *collective*, or *distributed* among the components: there is no central "leader" or "director" who decides what the other components should do. All molecules contribute equally to the maintenance of the organisation. A more complex example of such distributed organisation is found in the brain: the brain works as a whole; there is no part of the brain or neuron that tells the others what to do.

An important benefit of such distributed order is *robustness*: the system can take a beating and will not easily be brought off balance. This is because the system is *redundant*: if one component fails, there are enough similar components to take over. For example, if one magnet is forced out of its direction, it will be pulled back by the others. After damage to one part of the brain, for example when a tumour is removed, people remain able to lead a largely normal life, because the other parts of the brain replace the lost functionalities. On the other hand, damage to a computer or computer programme generally causes a complete shutdown. This is because a computer is not self-organising, but dependent on the detailed instructions of the programmer. If something goes wrong with that, the system cannot repair itself.

3.4 Nonlinearity

In mathematics, a function *f* is **linear** if, when applied to a sum, the result of the function is equal to the sum of the results for each of the components:

$$f(a+b) = f(a) = f(b)$$

It follows that:

$$f(2a) = f(a+a) = f(a) + f(a) = 2f(a)$$

More generally, this means that if a component is multiplied with any number *k*, the new result can be found by simply multiplying the old results with *k*:

$$f(k \cdot a) = k \cdot f(a)$$

Applied to concrete systems, linearity means that the result or effect [*f(a)*] of a process is *proportional* to the cause [*a*]. For example, if we push a car twice as hard (or if two people push instead of one), the car will move forward twice as fast. This means that small causes necessarily lead to small effects, and large causes to large effects.

In most quantitative models (such as in Newton's mechanics), we try to assume linearity as much as possible, because this makes the calculations easier. In practice however, most systems, and especially self-organising systems, are **nonlinear**. This means that the effect increases faster or slower than the cause.

In for example the magnet case, a small fluctuation (three or four little magnets that align themselves accidentally) can lead to a very large effect (the material becomes completely magnetic). Conversely, a large cause (for example a heavy shock that causes thousands of little magnets to change direction) can lead to an unnoticeable effect (because the remaining aligned magnets pull the others back to their place).

3.5 Chaos

The situation in which small causes have large effects is called *sensitive dependence on initial conditions*. Here, the initial conditions play the part of the cause, or the state of the system at the start of the process that is being examined. The system is called sensitive, because it reacts strongly to even the smallest variation in the initial conditions. In other words, the smallest, hardly visible fluctuation can lead to a completely different result.

<u>Examples:</u>

- A pencil standing upright on its tip will almost immediately fall to one side. The smallest disturbance (a breeze, an irregularity on the table surface, some more air molecules colliding with it from the left than from the right) is enough to break the balance and cause the pencil to tilt in a certain direction and fall. However, we can generally not predict in which direction it will fall.
- Another classical example is the famous *butterfly effect*. The system of equations that allows us to predict the weather, based on air pressure, temperature and wind direction, is to a great extent nonlinear. That is why the weather is strongly dependent on small fluctuations in the initial conditions, such as a bit more wind here or there. In principle,

> the fluttering of a butterfly in Tokyo may cause a hurricane in New York a few months later.

Systems of which the behaviour is so sensitively dependent on the smallest fluctuations are called *chaotic*. This means that they can suddenly change greatly due to a minimal influence, and thus behave in a very irregular, unpredictable manner. Examples of this are found in the turbulent currents at the bottom of a waterfall, or the weather, which can change from day to day and whose changes cannot be predicted more than a couple of days in advance, even with the most powerful computers.

The existence of this nonlinearity or **chaos** has an important implication for the Newtonian worldview: even systems that are in principle deterministic (i.e. their further evolution is completely determined by their original conditions) are in practice *unpredictable*, because we can never establish the initial condition with perfect accuracy, and because the smallest mistake can lead to very great deviations.

3.6 Far-from-equilibrium systems

In thermodynamics, systems are called **closed** if they do not exchange matter or energy with the outside world. Such systems evolve naturally to a state of equilibrium of maximal entropy or maximal disorder (the second law of thermodynamics, see 10.3).

Self-organisation implies a reduction of disorder and is therefore in principle impossible in a closed system. Consequently, self-organising systems need to at least lose entropy, which typically happens in the form of heat (energy with a high degree of entropy), which is released into the environment. Such release of entropy is called *dissipation*. Once all available energy has thus been released in the form of heat, the activity shuts down and thus the system reaches equilibrium. This is what happens with magnetisation as soon as all the little magnets have been aligned: there is no further movement.

Another way to enable self-organisation is to constantly add energy or matter with a low degree of entropy. (This can also be described as the input of "negentropy", i.e. negative entropy, in other words, order.) Through the normal processes within the system, this is converted into high entropy, which in turn needs to be dissipated. Such systems never reach a thermodynamic equilibrium, since they continue producing and dissipating entropy. This is why they are called "far from equilibrium". Such systems continue working and do not shut down. The dynamic organisation that characterises them is called a **dissipative structure**.

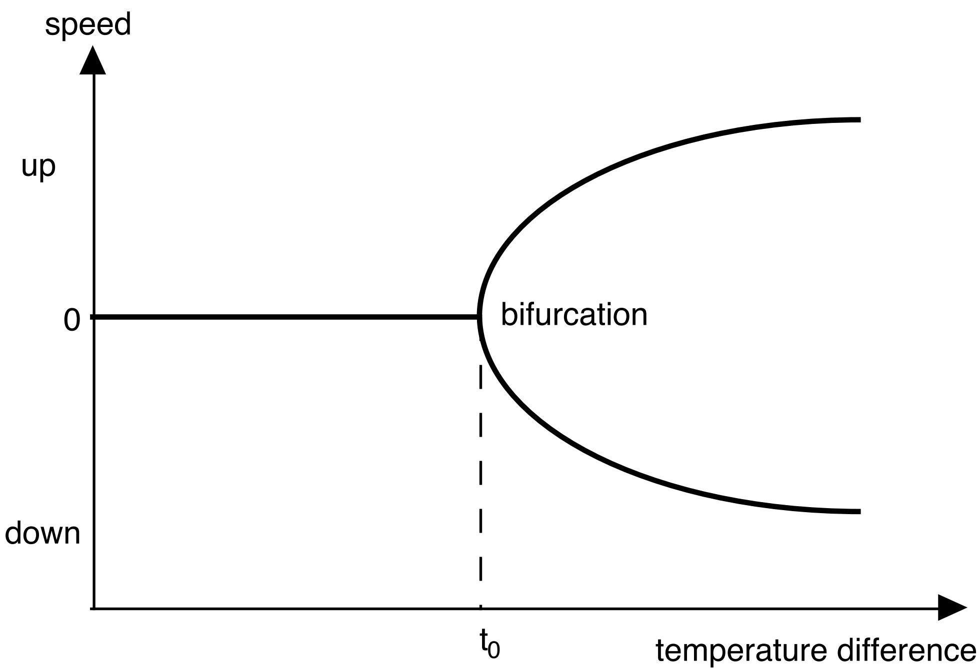
The Rayleigh-Bénard convection is the simplest example: the rolls or cells form a clear organisation or structure, which is not static, but rather based on the continuing, circulating flow of the fluid. Energy with low entropy is introduced through the heating at the bottom. This is released or dissipated in the form of high entropy at the cooler surface. Other examples are vortexes, flames, ocean currents (such as the Gulf Stream) and whirlwinds: these are dynamic structures that can only be perpetuated through the introduction of energy (solar heat, fuel, or running water). Even more important, all living beings are dissipative structures: they can after all only survive through a constant intake of matter and energy in the form of food, which is subsequently excreted in the form of waste and heat.

Far-from-equilibrium systems are nonlinear: the input of energy strengthens certain effects, while the loss (dissipation) of energy weakens other effects. This makes the system generally unpredictable or chaotic: the further from equilibrium (the more energy is injected), the more chaotic. This can easily be observed with a tap: the wider the tap is open, the faster the water is pushed out. Thus the flow has more energy, but the current also becomes more irregular or turbulent.

3.7 Bifurcations*

For nonlinear systems, there generally is more than one possibly solution to the system of equations that describes the system. Thus, the system can end up in more than one stable situation or "steady state". The further from equilibrium, the more solutions. The appearance of such choices when the system gets further away from equilibrium is called a *bifurcation*.

Example: Bénard rolls: Initially, when the difference in temperature between the top and the bottom is still small (close to equilibrium), the fluid remains immobile (speed = 0). When the difference in temperature increases up to a certain value t_0, the rolls start to appear. For a given roll there are however two possible rotation directions, namely clockwise and anticlockwise. That is to say that on a given location, the fluid has to "decide" whether to start moving up or down. The system has no *a priori* preference, but it has to choose one of the two alternatives. The final "choice" depends on accidental factors. These are generally too small to observe, which is why the result (the path that the system chooses when it reaches the bifurcation) is unpredictable.



The further from equilibrium the system gets, the more bifurcations it undergoes and thus the more choices it gets. Complete chaos arises when there is an infinite number of choices, and the system jumps continuously and unpredictably from one option to the next. Such turbulence arises when the heat under the Bénard rolls continues to rise, until the fluid starts boiling.

Chapter 4. Systems theory

4.1 Holism and emergence

A fundamental problem of the theories of both self-organisation and evolution is that they do not actually explain what organisation and complexity really are. The reason for this shortcoming is that they implicitly still start from the reductionist approach, according to which every phenomenon can be reduced to its smallest material components: elementary particles. These particles are by definition not complex and therefore show no organisation. The organisation that we observe then appears to be no more than a superficial phenomenon, for which there is no place in the Newtonian worldview.

In order to overcome this problem, we will now introduce two closely related approaches, systems theory and cybernetics, which have both examined organisation, albeit from complementary angles. Against reductionism, systems theory places **holism**, which emphasizes the *whole* rather than the *parts*. In other words, the cohesion is more important than the individual components.

The simplest way to express this is the well-known expression: *the whole is more than the sum of its parts*. However, this begs the question, of *what* is there more? The answer is **emergence**. Emergent properties are properties of a whole that cannot be reduced to properties of the parts. This can best be explained by means of several examples.

<u>Examples:</u>

- A car as a whole has the property that it can drive. None of the individual parts, such as the body, wheels, motor or axes, can drive on their own. The ability to drive is therefore an emergent property. The weight of the car on the other hand is not an emergent property, since it is not more than the sum of the weights of the parts.
- Following the same principle, an animal has the property of being “alive”. The individual molecules comprising the animal are however dead matter.
- Table salt has the properties of forming crystals, being edible and having a salty taste. Table salt (known in physics as sodium chloride) is a compound of the elements sodium (a highly reactive metal) and chlorine (a poisonous gas). Neither of these parts has the properties of table salt.
- A piece of music has the properties of rhythm, harmony and melody. The individual notes comprising the piece of music however do not have these properties.

Self-organisation is typically characterised by the appearance of emergent properties. As we have seen, Bénard rolls are characterised by their rotation direction: clockwise or anticlockwise. These rolls however, consist of molecules that move independently from each other in a straight line until they collide. Therefore, the molecules do not have a rotation direction.

4.2 What is systems theory?

The ideas of emergence and holism were first formulated around 1925 in the work of authors such as Smuts and Whitehead. The ideas remained however rather vague, with a mystical tint, and therefore not very scientific. Thereafter, the biologist Ludwig von Bertalanffy introduced the idea of a system as a coherent whole that in principle could be described in a precise, mathematical form. In 1955 he created, together with scientists such as Boulding and Rapoport, the *Society for General Systems Research* (still extant as the *International Society for Systems Science*, or ISSS). This society aimed to develop a *general* systems theory, that is, a theory that would be applicable to all possible types of systems, whether they are physical, living, social or intellectual systems. Thus, general systems theory would lead to a unification of science, which until then had been fragmented in a manifold of disciplines: physics, chemistry, biology, psychology, sociology etc.

To understand this, we have to examine the underlying philosophy of systems theory. A **system** is defined as a collection of components, connected through relationships. These components may be of very diverse nature: atoms, molecules, cells, transistors, neurons, people, companies, symbols, concepts, etc. Relationships represent the influence of one component on the behaviour of another. For example, an atom exerts an electromagnetic force on another atom, a manager gives orders to their subordinate, a neuron gives an electrical pulse to another neuron. Moreover, the system as a whole has its own *identity*, which distinguishes it from its environment or background. Examples of systems are an organism, organisation, planet, computer, bicycle, number system, or the periodic table of elements in physics.

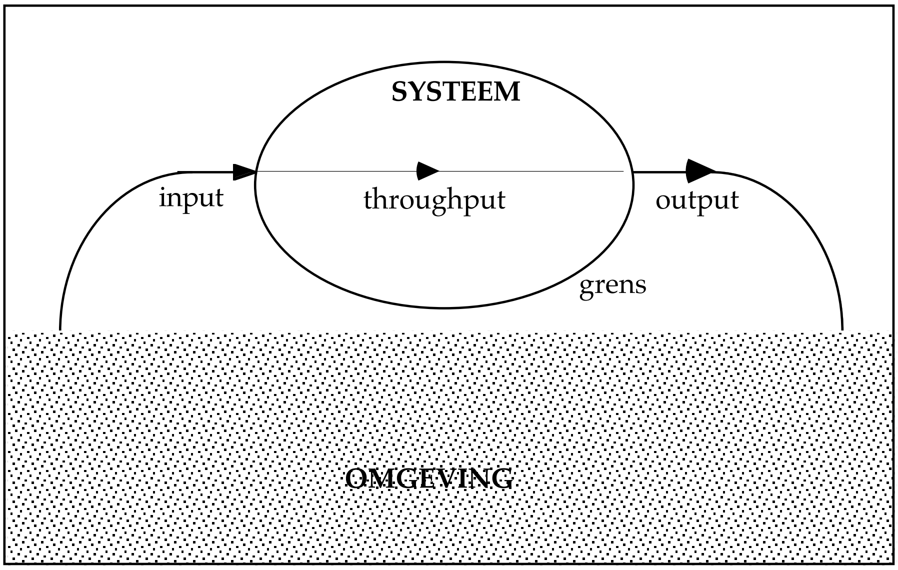
A general system is intrinsically abstract, independent of the concrete matter that makes up its components. This means that physically very different systems can be *isomorphic*, i.e. they can have a similar structure or organisation. For example, a society is in certain respects isomorphic to an organism, and a computer behaves in certain respects like the brain. Certain analogies or isomorphisms enable us to better understand complex systems, such as the brain. Since the organisation is defined in an abstract way, mathematical models of systems are in principle possible.

Because of this parallelism between systems with components of different kinds, it now becomes possible to rise above the barriers between disciplines: after all, by making the material or concrete components abstract, we can now understand the phenomena (for example life, thought, societies...) as different instantiations or actualisations of the same fundamental organisation.

General systems theory is actually not a theory in the strict sense, but rather a way of thinking. It is therefore often denoted with terms such as systems approach, systems science, or systems research. We can also see it as an attempt to create a universal language, which would allow us to describe all possible systems, or as a conceptual framework with very broad applications, among which are system analysis, problem solving, system design (for example technological systems or organisations), and the integration of data from different disciplines.

4.3 Basic components of a system

The insight that was fundamental to von Bertalanffy's approach is that real systems, such as living beings, are always **open**: they exchange matter, energy and/or information with their environment. Cut an organism completely off from its environment (for example in an airtight box) and it will soon break die because of lack of oxygen, hunger and thirst. In the Newtonian worldview on the other hand, it was implicitly assumed that the examined systems (such as the solar system) were **closed**: all matter and energy that influence the system are already part of the system and will stay there; it is not necessary to take the environment into consideration.



The step from closed to open systems requires the introduction of several new fundamental concepts. The system is separated from its environment by a **boundary**. Whatever is inside this boundary is by definition part of the system; that which is outside the boundary is part of the **environment**. For example, for a human being, the skin is the boundary between system and environment. Moreover, every system interacts across this boundary with its environment through the exchange of matter, energy and/or information. Incoming interactions are called **input**; outgoing interactions are called **output**.

Normally, the input is processed in some way inside the system, and the result of this processing forms the basis of the output. The system can thus be seen as a process that transforms input into output. Such "passing through function" is sometimes called *throughput*.

<u>Examples</u> of input-output systems:

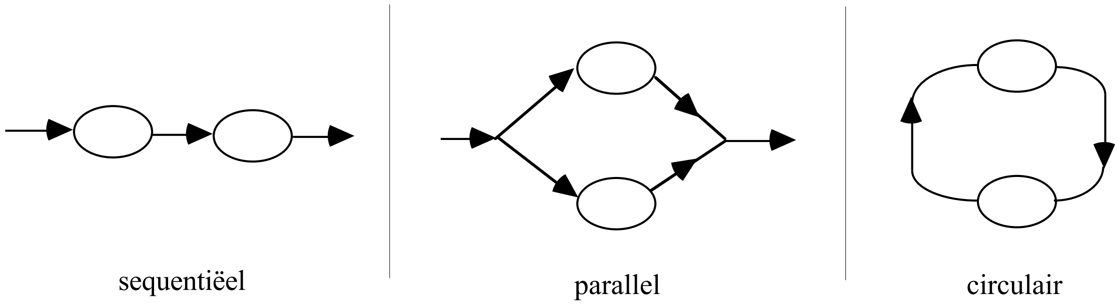
- The *body*: food, drink, oxygen = input; faeces, urine, CO_2 = output; digestion and the burning of calories = throughput
- The *brain*: perceptions, stimuli = input; decisions, actions = output; processing of received information = throughput
- A *computer*: movements across keyboard and mouse = input; information produced on screen or in a printer = output

4.4 Coupled systems

Input and output make it possible to *couple* different systems, by using (part of) the output of one system as input for another system. For example, the berries of a plant (output) are eaten by birds (input), while bird droppings produce minerals (output) that are taken in by plants (input). Coupled systems are to a certain extent dependent on each other.

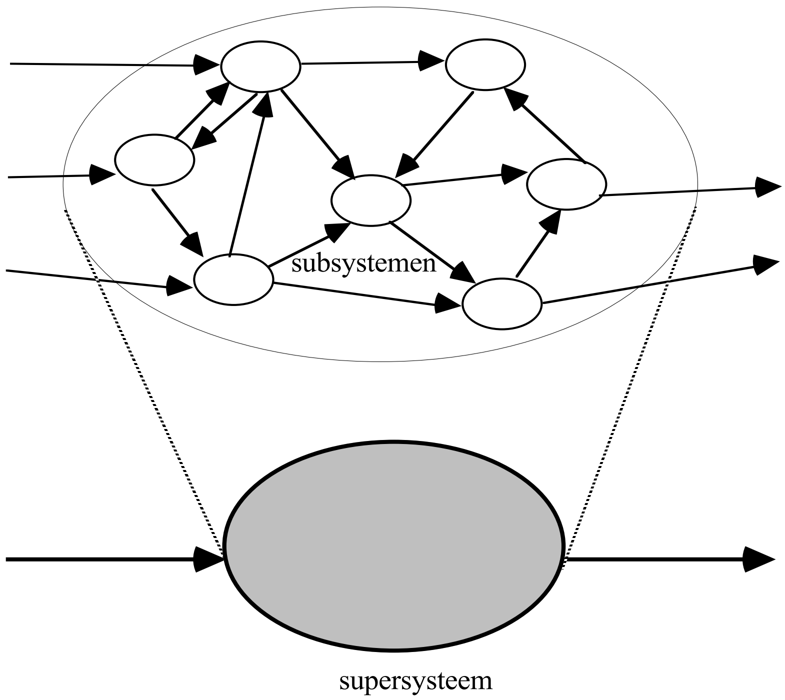
There are three elementary forms of coupling:

- **sequential** (or serial): the output of the first system forms the input of the second; the systems follow each other in unambiguous order
- **parallel**: both systems receive input from the same source (an unspecified third system) and deliver output to the same destination; they work side by side, in parallel
- **circular**: the input of the first system forms the output of the second and vice versa; they "feed" each other and together they form a loop; this is also called *feedback* and will be discussed in more detail in the next chapter.



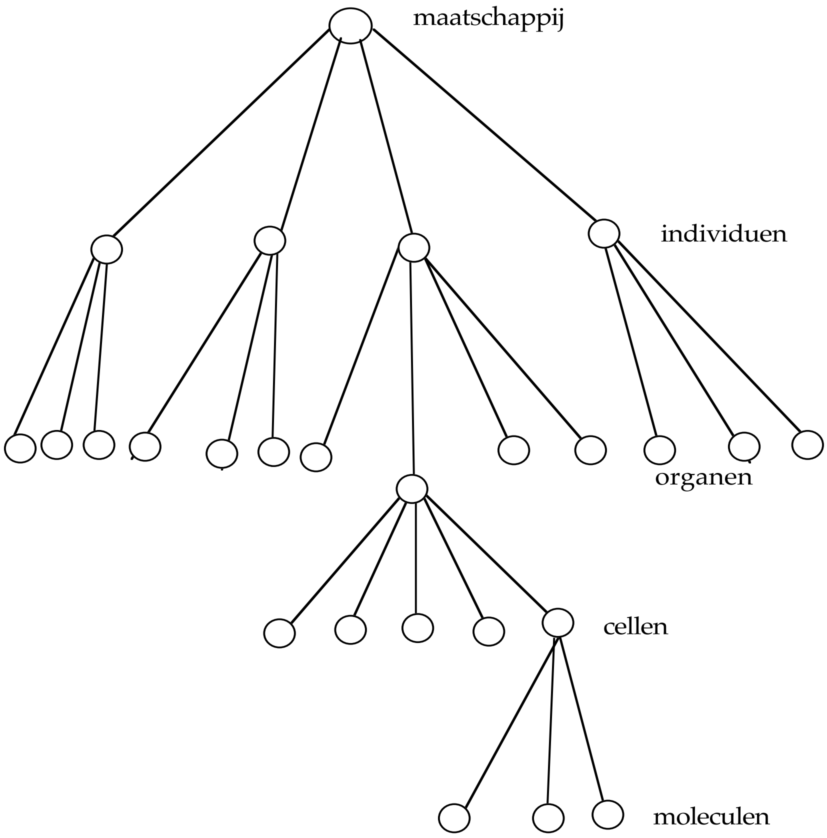
4.5 Subsystems and supersystems

When we are dealing with a more complex system, consisting of several systems coupled in different ways, we speak of a *network*. When the systems in the network form a coherent whole, which can be clearly demarcated from its environment, they define a larger, encompassing system: a **supersystem**.



The illustration shows a network of coupled systems (above), each with their input and output. This complex whole is presented in a simplified way (below) as a single, overarching system, the supersystem, which has its own input and output. In such representations, the supersystem behaves like a so-called *black box*: it is impossible to look inside and see the internal structure or components, but it can be studied by observing the relation between input and output.

The components or parts of a system are in turn generally systems themselves: **subsystems**. Normally each system contains subsystems, while it is encompassed by one or more supersystems. For example, for the system “human”, society is a supersystem, while the heart and the brain are subsystems. These organs themselves consist of cells, which are subsystems of the subsystem. The cells have molecules as subsystems. Such supersystems and subsystems together form a **hierarchy**, where each system on a certain level of the hierarchy consists of other systems from the lower level, as illustrated below.



This hierarchical organisation of complex systems implies that it is always possible to “zoom in” (enlarge, for example with a microscope) in order to study smaller subsystems, meaning that you go down in the hierarchy. You can also “zoom out” (view from a greater distance, by means as it were of a “macroscope”, as suggested by J. de Rosnay) to better observe the larger supersystems. This means that you go up in the hierarchy. Both are movements in the “scale dimension”. This means that you continuously change the scale of the representation, like a satellite that takes photos on a smaller scale (e.g. of a continent) or on a larger scale (e.g. of a city).

Chapter 5. Cybernetics

5.1 Goal-directedness

Cybernetics developed around the same time as systems theory, with similar intentions to arrive at a universal theory of organisation that would apply to all disciplines. Some of the most famous founders of the approach are Norbert Wiener, W. Ross Ashby, Heinz von Foerster and Gregory Bateson. Cyberneticists and systems theorists collaborated and strongly influenced each other. The difference is that systems theory focuses mainly on the **structure** of a system ("how does it work?"), whereas cybernetics focuses mainly on the **function** ("what is it for?").

The function of a system or subsystem can be defined as the **goal** that the system aims to realise and for which it has been created. In practice, we see that many systems have a purpose: human beings, automatons, organisations, animals, and in fact all living beings. For example, plants strive to absorb as much light as possible, and will therefore make their leaves grow in the direction of the sun. Lions aim to catch prey, and will thus adjust their actions (for example stalk the prey, run towards it, pounce it, bite it to death) towards this goal. The subsystems of the system "lion" (for example teeth, claws, eyes) contribute to the attainment of that goal and therefore each has its own function.

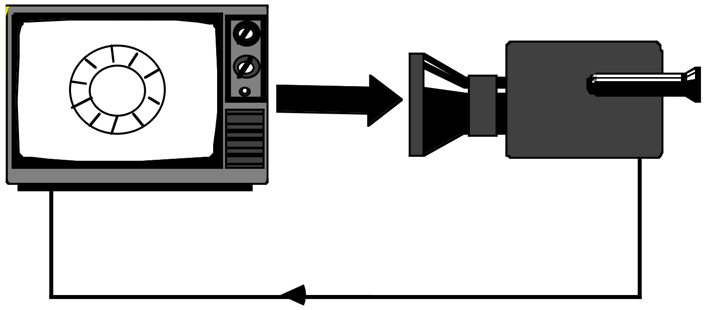
With goal-directed actions, we assume that the goal will not necessarily be reached, and if it would be reached, that we will generally not know when it will be reached. In the Newtonian worldview, however, there is no room for "trying", "succeeding" or "failing", because the outcome of each action is always predetermined. Cybernetics does not considers this a problem, because it assumes, like the theory of evolution and other theories before it, that we cannot predict the future anyway, and therefore have to rely on trial and error.

A more fundamental problem is that a goal is by definition in the *future*, but still determines the behaviour in the present. This contradicts Newtonian mechanics, which assumes after all that the effects are already fully determined by their initial state or cause, which is by definition in the *past*. (The religious worldview has no *a priori* problem with purpose, since it assumes that God himself is purposeful and has to a certain extent imposed his goals on the world.) Cybernetics has solved this paradox by introducing *circular* causality: feedback.

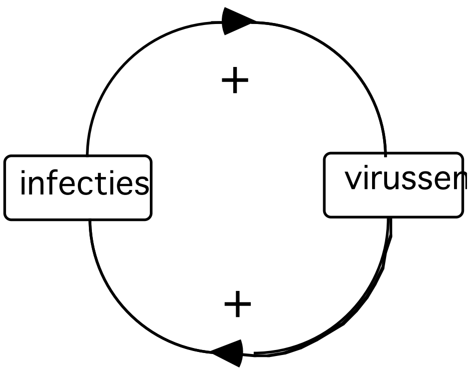
5.2 Feedback

Feedback can be defined as a process in which the effect leads back to—or affects—its own cause. In terms of systems theory, it can be viewed as a situation in which the output of a system is re-entered (directly, or via one or more intermediary systems) into the same system, so that it simultaneously plays the part of the cause. Circular interconnection is an example of this.

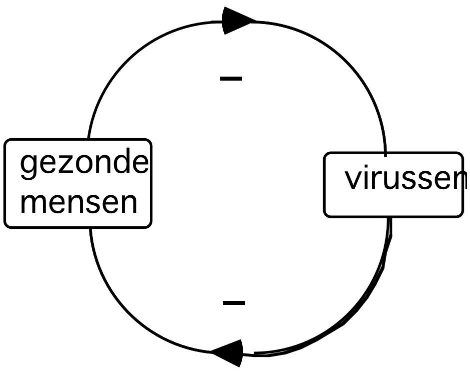
<u>Illustration:</u> Imagine a video camera that registers an image (cause, input) and sends the image on to a screen on which the image is projected (effect, output). Now imagine that this video camera is pointed at the screen. In this situation, the image on the screen is both cause and effect of the process. In practice, this image will show all sorts of bizarre, abstract patterns, which can be understood as the result of a kind of self-organisation.



There are two particular, fundamental cases of feedback: positive and negative. With **positive feedback**, the cause will *reinforce* itself. Let me illustrate this with an example: the feedback relationship between humans with a cold and the spread of viruses. The more people have been infected (cause), the more viruses will be spread through sneezing and coughing (effect). The more viruses are spread, the more people will be infected (feedback to the cause). Thus, more infections lead, via more viruses, again to more infections. The total amount of infected people will thus continue to increase, until everyone who was susceptible to infection has been infected.

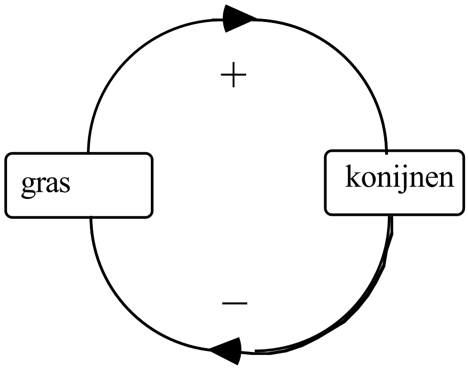


Such positive feedback *increases the deviation* from the initial state (nobody infected). This leads to an explosive, ever-faster growth, which will only end when all available "resources" for the process (in this case uninfected people) have been exhausted. Other examples of this are the chain reactions leading to nuclear explosions, snowballing phenomena (such as the increasing debt, which leads to higher interest payments and thus further increasing debt), or vicious circles (such as poverty which leads to deficient education, leading to low-paid jobs or unemployment, and thus more poverty). Positive feedback is generally the basis of what we have called sensitive dependence on initial conditions, and can be seen as the aspect of nonlinearity that produces large effects from small causes.



In the diagram, a positive causal influence (more → more, or potentially: less → less) is denoted by an arrow accompanied by a + sign. The two circularly connected arrows both have a positive sign, which means that the feedback loop as a whole is positive. The same result would be achieved with two negative arrows, which after all negate each other's influence (two negatives make a positive). Moreover, if we would replace the number of infections with the number of healthy (as yet uninfected) people, the influence of the viruses on this would be negative, while the number of healthy people would have a negative influence on the number of viruses. The net effect remains the same positive feedback: more viruses leads to less healthy people, and therefore to more viruses. In general, a circular connection is positive if there is an even number of minus signs (or if there are none).

The opposite situation, where a cause weakens itself and deviations decrease, is called **negative feedback**. This is the case for a circular connection with one (or an uneven number of) negative influence(s). An example is the relationship between rabbits and grass. If there is more grass, there is more food for the rabbits, so the number of rabbits increases. This is a positive influence, as in the previous example. The influence in the opposite direction is negative, however: more rabbits eat more grass, so that less grass remains. This in turn reduces the number of rabbits, so that less grass is eaten and the grass gets a chance to grow.



Every deviation (whether more or less) thus leads to its opposite (less or more). Negative feedback can be seen as a nonlinear mechanism that reduces large causes to small effects. In general, this leads to a *stable* equilibrium, since every deviation of the equilibrium is reversed and the system returns to its equilibrium. Such equilibrium can be seen as a kind of implicit "goal" to which the system always returns. Another example can be found in the market mechanism: if a product becomes scarce, the price will increase (negative influence of supply on price). Producers will then receive more money and will therefore bring more of this product onto the market (positive influence of price on supply), which will bring down the price.

Note*: negative feedback does not always lead to equilibrium. If there is a *delay*, i.e. the suppression of the deviation comes much later than the deviation itself, the deviation has time to grow, causing the correction to be too great, so that a deviation will now arise in the other direction. This new deviation will in turn be delayed in its return to the other side, and so on. The result is oscillation or a cyclical movement around the equilibrium. This is an example of a limit cycle (see 11.6).

5.3 Control

The core concept of cybernetics is what we will call **control**. Other names for this concepts are *regulation*, *steering*, or *direction*. A control system is a goal-directed system that tries to keep deviations of its goal under control through suppression of all disturbances. This goal is a state or a set of states that the system somehow prefers and which it therefore will try to attain if it is not there yet, and stay there otherwise. Control is an *active* form of negative feedback. This means that the target state (or goal state) in itself is not an equilibrium that the system spontaneously reaches (like a ball that rolls spontaneously into a put and remains there), but a "far-from-equilibrium" state (see 3.6). The system has to actively intervene and expend energy to reach and maintain it.

Thus, the system has to undertake **actions** to counter every deviation from the goal state. In cybernetics, this is called "**compensation** of **perturbations**": the perturbations are all obstacles, problems, fluctuations or disturbances that cause the system to deviate from its goal. These are caused by all sorts of uncontrollable or unpredictable factors, such as e.g. accidents, interferences, conflicts, mistakes, or weather changes. They have to be neutralised or compensated for with the correct counter-actions.

Example: If I try to reach the jetty on the other side of the river with my motorboat, I will be confronted with all sorts of perturbations, which make my boat diverge from the goal: gusts of wind, currents, dangerous rocks, etc. To reach my goal, I will have to keep adjusting my steering. If for example the wind blows my boat too much to the left, I will have to turn the rudder to the right to compensate for this disturbance. If after that a sudden current drags me to the right, I will have to steer left to remain on course. Without these continuous, active corrections, the boat will never reach its goal, but drift around uncontrolled, prey to wind and weather. It is this metaphor of the steersman (*kybernetes* in Greek) that has given cybernetics its name.

What else is needed to efficiently steer towards the goal?

- **Perception**: to suppress perturbations, you first have to perceive them; "senses" or sensors are therefore needed. These sensors supply information about the state of the system and its environment, and this information has to be interpreted or processed so that the precise deviation can be determined.
- **Knowledge**: to decide which action is adequate for which deviation, knowledge is needed. For steering a motorboat, this is rather simple. However, if I would want to steer a sailboat, a car or a plane, I would have to undergo a lengthy learning process.

5.4 An example: the thermostat

Let me clarify the concept of control system with a very simple, concrete example: the thermostat that keeps our house nice and warm, whatever the weather. The thermostat is the classical prototype of a purposeful, cybernetic system. The operation is very simple: a sensor in the thermostat reacts to changes in temperature. If the temperature is higher than a certain value that has been set by users (e.g. 21º C), nothing happens. However, if the temperature falls below that value, the thermostat turns on the heating system (for example by closing an electrical contact). This warms up the room and raises the temperature. As soon as the temperature is higher than 21º, the thermostat turns the heating off, so that the room can cool down again. When this cooling down causes the temperature to get below 21º, the heating is turned on again and the cycle starts anew. The net result is that the room temperature remains around the set temperature and fluctuates very little, whether it is warm or cold outside. The thermostat regulates or controls the temperature of the room and makes it independent of the outside temperature.

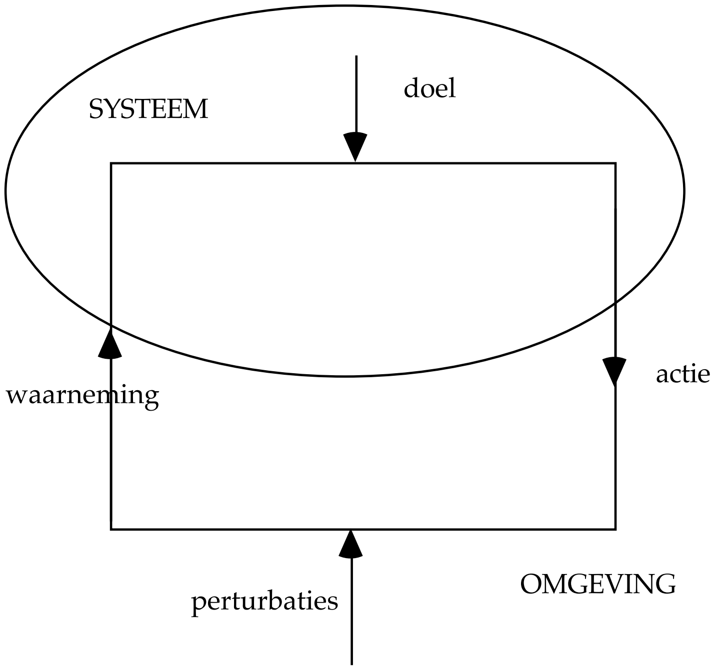
Let us discuss the core cybernetic concepts in this example:

- the **goal** (or function) of the thermostat is to reach and maintain the set temperature
- the **perturbations** or disturbances are the changes in the outside temperature that threaten to make the inside temperature deviate from the goal
- the **perception** is the inside temperature as registered by the sensor
- the **action** of the thermostat consists in turning the heating on or off
- the **knowledge** of the thermostat consists in knowing that it should turn on the heating if the temperature is lower than the goal, and otherwise turn it off (imagine that the thermostat would do the opposite, continue heating when it is already warm enough, and let it cool further when it is too cold!)
- the **feedback** is that the result (e.g. rising of the temperature) of the action (output) is observed again by the thermostat (input).

This feedback is **negative**, because the observed deviation is compensated by an action in the *opposite* direction (for example by turning off the heating when it becomes too warm). If the action would go in the same direction, as we suggested above, the feedback would be **positive**, and the people in the room would either freeze, or get too hot.

5.5 Core components of a control system

We will now generalise the example of the thermostat and discuss the fundamental components of an arbitrary goal-directed system using the figure below. This diagram is a special case of the elementary input-output diagram in section 4.3, with perception as input and action as output, but with the addition of a goal, and the feedback from action to perception. Because action leads again to perception, the diagram is a closed loop. Because perception does not only depend on action, but also on what is happening in the environment, we need to indicate an additional input in the loop, namely the perturbations.



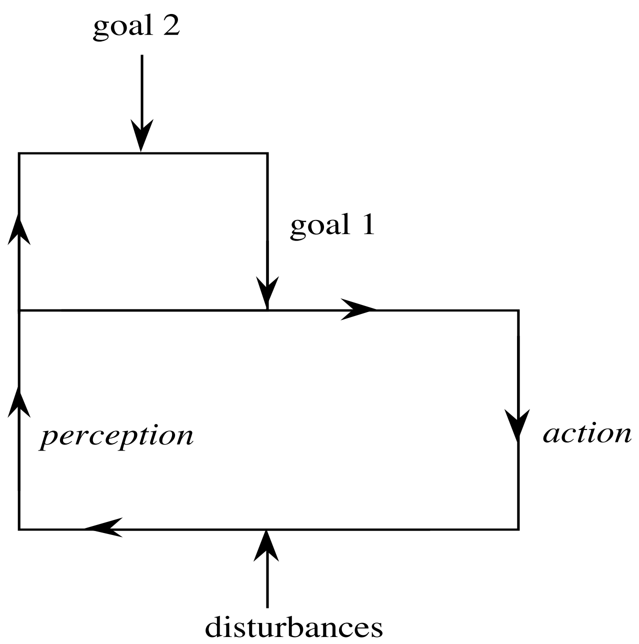
Example:

To illustrate that this diagram is perfectly universal, we will now discuss a far more complex example: a multinational company. Such a company is a control system with the goal of making as much profit as possible in the long term. The company has at its disposal a large variety of actions to reach this goal: starting publicity campaigns, recruiting managers, dismissing redundant staff, automating production, lobbying politicians, etc. The perturbations are all the events in the environment over which the company has no direct control, but which influence the realisation of its goal: activities of competitors, demands from customers, the fluctuating economical climate, natural disasters that impede transport, decisions of governments to levy taxes, fluctuations in fuel prices, etc. Actions and perturbations together determine how successful the company is in the attainment of its goals. To estimate (perceive) this, the company needs to gather as much information as possible, about a variety of factors: market share, sales, production costs, brand awareness, reputation among the public, motivation of staff, etc. Based on this information, the company will decide to adapt its actions to the circumstances, and for example halt dismissals, because it observed that this damages staff morale and reputation too much, while it contributes little to cost reduction.

5.6 Control hierarchy

The diagram above is too simplistic in the sense that it does not account for the fact that a system tends to have several goals, of which some are *subordinate* to others. Subordinate goals or sub-goals are goals that are of themselves not very important, and are only important because they contribute to the attainment of a higher or more general goal. For a lion for example, the most general goal is to survive and reproduce. Subordinate to this is the goal to still its hunger. On a lower level this leads to the goal to hunt prey, for example a zebra. Subordinate to this are the goals to run to the zebra, pounce it, bite it in the neck, etc. Thus goals and sub-goals form a complex hierarchy of subordination or dependency.

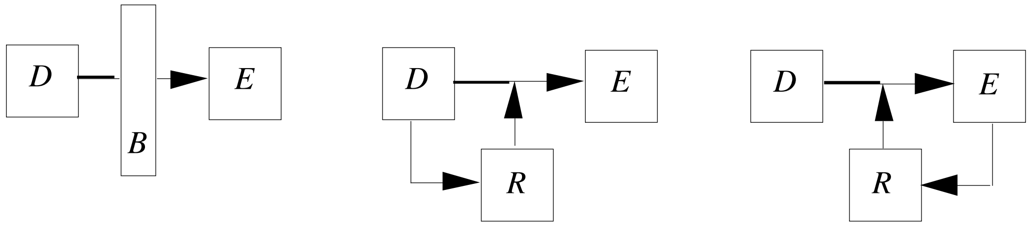
The diagram above can easily be extended to illustrate this hierarchy. It suffices to add a second control loop on top of the first one, so that the first, subordinate goal in fact becomes the result of the action of the second, higher goal. This means that the higher order control loop sets the goals of the subordinate goals as its action, depending on the perceived circumstances. Again, this can be best illustrated with the example of a thermostat.



The goal of a thermostat is to reach and maintain a certain temperature *T*. Our overarching goal however is to keep the temperature comfortable for the inhabitants of the house, but without wasting energy. For this, we can build a higher order control system that controls the thermostat by for example varying the goal temperature *T* depending on occupation. For example, if there are people in the room, set *T* to 21º. When the room is empty, set *T* to 16º to conserve energy. An infrared sensor can check if someone enters the room. If the sensor detects people, the goal temperature is increased. As soon as they leave the room again, the goal temperature decreases. Such a control system can regulate several rooms at the same time, and set the thermostat for each room separately. This control system is a "meta system" (see 14.2) with two levels: control of the temperature within a room, and control of the goal temperature across all rooms.

5.7 The three fundamental control mechanisms

Thus far, we have only discussed control through (negative) **feedback**. Although this mechanism is the most essential, there are two other important control mechanisms: **buffering** and **feedforward**. Let me list the advantages and disadvantages of each. The illustration below shows that these mechanisms are special cases of sequential, parallel and circular connections, respectively. In this illustration, D stands for disturbance (perturbation), B for buffer, R for reaction and E for the "essential variables" that define the goal. The width of the arrows indicates the size of the influence: the task of the control system is to make the influence of D on E as small as possible.



- Buffering

Disturbances D are passively absorbed or dampened by a "buffer" *B*. There is thus no active intervention. The advantage of this is that the control system does not waste energy for the preparation and implementation of compensating actions.

Examples of buffers:

- shock absorbers and bumpers in a car to intercept shocks
- a well-insulated wall will protect the room from temperature fluctuations. This eases the work of the heating system
- a water reservoir reduces fluctuations (e.g. dry periods) in the amount of available water

Disadvantage: buffering can lessen the effect of uncoordinated fluctuations, but it cannot systematically drive the system to a goal that is itself not an equilibrium. For example, an insulated wall alone cannot keep the room warm: input of heat (in the form of fuel or electricity) is required for heating, since otherwise equilibrium would be reached when the inside and outside temperature are the same. Shock absorbers alone cannot keep the car on track: that requires a motor and a driver.

- Feedforward

Disturbances D are compensated for by reactions *R before* they can influence the goal *E*. The control system thus **anticipates** the effect of the perturbations on the goal. This is important in situations where time is needed to implement the necessary actions: by starting to compensate in advance, the control system prevents that the deviation would last too long or become too large. Feedforward is crucial for deviations that need to be avoided at all costs, for example because they could destroy the system.

Examples:

- A thermostat with an *external* temperature sensor can turn on the heating when it gets cold outside, before the room gets cold.
- If a multinational company learns that a competitor is preparing a large publicity campaign, they will anticipate that this may lead to a loss in market share, which is difficult to recover from. To forestall this problem, they themselves can start a campaign as soon as possible.
- A zebra cannot afford to be jumped on by a lion, since this attack might well be fatal. Thus, the zebra had better anticipate the attack and flee when the lion is still at a safe distance.

Disadvantage: it is not always possible to anticipate exactly how the perceived perturbations will influence the goal. For example, if someone turns on the oven in the room, this will be enough to counteract the cooling down outside and the heating will not be required. The rumour about the new publicity campaign could be unfounded and the lion may not be hungry. Such mistakes in anticipation do not only lead to a waste of energy, but possibly also to overcompensation, such as a room that becomes too warm.

- Feedback

Disturbances D are only compensated by actions *R after* they have caused a deviation from goal *E*. The advantage is that we can now be certain of exactly which deviation is happening. We do not need to know what has caused the deviation in order to anticipate possible consequences. The only thing we need to know is which action can compensate for this deviation. This simplifies the control problem.

Examples:

- The thermostat will normally only turn on the heating after the temperature has sunk below the goal temperature, so that there is no longer any doubt about the need for extra heating.
- The central bank of a country will watch the economic growth closely. If this deviates too much from the goal, it will intervene by either increasing the interest rate, which slows growth, or decreasing the rate, which stimulates growth. Because of that, it is not necessary to know the very complex and hard-to-predict factors that influence economic growth.

Disadvantage: a deviation from the goal state has to occur before the disturbance can be suppressed. Thus, feedback control is never perfect: there will always be deviations. By acting as soon as possible, we can however aim to keep those as small as possible.

5.8 Knowledge

To actively suppress disturbances, a system has to:

- be able to implement a sufficiently large variety (see 9.2) of actions to handle every perturbation. This is Ashby's *law of requisite variety*.
- know which action is appropriate for a given perturbation. This is the law of requisite knowledge.

This means that the larger the variety of perturbations that confront the system, the larger the repertoire of required actions, but also the more knowledge the system needs. Complex, variable environments thus require highly developed systems, which can handle wide-ranging problems.

At the most fundamental level, **knowledge** is expressed in the form of a **condition-action rule**: IF a certain perturbation is observed, THEN implement certain adapted action. In short: IF condition, THEN action. Or even shorter:

condition → action

Example: for the thermostat the following rules apply:

temperature too low (lower than goal temperature) → turn heating on

temperature high enough → turn heating off

For the thermostat controlled by the presence sensor:

someone in the room → set goal temperature to 21º

room empty → set goal temperature to 16º

If the system uses feedforward, knowledge will also be expressed in the form of *predictions*, where a condition that is not a deviation entails another condition, which could possibly be a deviation, and to which the system can react in anticipation.

condition 1 → condition 2

<u>Example:</u> thermostat with external temperature sensor

outside temperature falls → inside temperature will fall

In combination with the first rule of the thermostat, this implies:

outside temperature falls → turn heating on

A system of prediction rules is called a **model**. The model represents processes in the environment in an abstract, simplified way. The model allows the system to *anticipate* perturbations or events. The model is however not an objective reflection of reality. It is a subjective representation of aspects that directly concern the system, and it serves to allow a goal-directed system to reach its personal goals.

Chapter 6. Complex Adaptive Systems

6.1 Background

In the early '80s, the Santa Fe Institute for Complex Systems was founded in New Mexico. Its goal was to achieve the interdisciplinary study of complex systems, by getting scientists from different centres to work together around this theme. Some of the most famous scientists that have been connected to the SFI are John Holland, Stuart Kauffman, Brian Arthur and Chris Langton. The ideas of these scientists have meanwhile become more and more popular in the scientific community. This has led to a study area that is usually called **complex adaptive systems** (CAS), and more recently complexity science(s).

This approach is related to and influenced by previous approaches, although it is conceptually less developed. Some of important new concepts of the CAS approach, such as "self-organised criticality", "percolation" or "NK networks" are therefore more complicated and have less obvious applications than those of systems theory and cybernetics. I will therefore not discuss them in this introductory book. As a matter of fact, many of these concepts go out of fashion after a couple of years of popularity, when it becomes clear that they are not quite as generally useful as previously thought—in contrast to concepts such as emergence, fitness and feedback, which have become part of general scientific thought.

The reason for the popularity of the CAS approach is mostly its innovative method to study complex systems: the use of *computer simulations*. This enables us to study more complex, less idealised systems. At the time of previous approaches, computers were after all not yet powerful enough to simulate really complex systems, so that most results were reached through purely theoretical reasoning.

6.2 Agents

The computer models of the CAS approach are generally based on the concept of an **agent**, as the elementary, active component of the system that is being modelled. Examples of agents are cells in a body, individuals in a society, birds in a flock, buyers and sellers on the stock market, ants in an anthill, autonomous robots in a lab, or molecules in a chemical reaction. Although there is no universally accepted definition of an agent, the term usually means an elementary, goal-directed system, which can undertake different actions. The goal can be explicit or implicit, but it usually boils down to the agent trying to maximise its "benefit", "success" or "fitness". When several agents pursue the same goal, they will usually enter into competition, but they are potentially able to cooperate.

In response to perceptions (incoming information), the agent will implement certain actions. For this, it follows simple *condition-action rules*. For example, IF the agent finds a piece of food, THEN it will eat it. Agents can sometimes learn or evolve individually and thus adapt their behavioural rules. For example, if food with a blue colour turns out to taste bad, the rule will be adapted, so that the agent will only try to eat non-blue food.

Their actions also influence other agents, which will react with actions of their own. For example, if an agent moves towards a piece of food, but sees that another agent has been ahead of him and has already eaten the food, it will change direction to a piece a bit further away. This may attract the attention of a third agent, so that it will also start moving towards the piece. In this way, an event or action somewhere in the system (e.g. a piece of food that appears or is eaten) will get several agents in motion, which will in turn incite others to action. This activity can continue without limit, or settle down in an equilibrium configuration.

6.3 What is a complex adaptive system?

A **complex adaptive system** (CAS) is a collective consisting of a large number of interacting agents. (More recently, the term *multi-agent system* (MAS) has also been used.) Collective patterns emerge from individual actions, and the system itself is self-organising. The entire system is also adaptive: a change in the environment will lead to a change in the system. The system *adapts* to new circumstances, and will generally reach equilibrium with them, even if the individual agents are not necessarily adaptive. A typical system of this type exhibits to various degrees the most important mechanisms that we have discussed in earlier chapters, such as **variation**, **natural selection**, **non-linearity**, **self-organisation**, **feedback**, **anticipation** and **hierarchy**, but in a complex tangle of interactions of which the consequences are generally hard to predict.

<u>Examples</u> of such CAS:

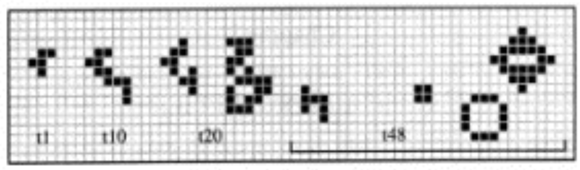
- *Markets*: depending on the prices set by other agents, agents can decide to buy or sell. This in turn influences prices and thus the behaviour of other agents. This can lead to collective phenomena, such as equilibrium, inflation, or speculation.
- *Self-organising chemical reactions*: the agents here are the interacting molecules that can self-organise to form dissipative structures.
- *Historical civilisations*: individuals interact with each other and with their environment (crops, cattle, ...). This can lead to growth or on the other hand, to famine and the disappearance of the society (e.g. the collapse of the Maya empire).
- *Ecosystems*: the actions of certain plants or animals influence others, which can lead to emergent phenomena such as symbiosis, parasitism, extinction, etc.

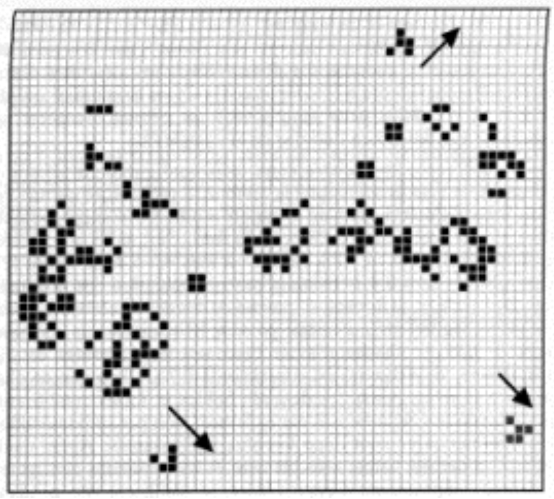
Computer simulations allow us to better understand what is happening—or could happen—with such systems under different conditions. The assumptions behind these simulations (rules of behaviour) are however rather simplified and artificial. Therefore they offer no guarantee that the results of the simulation will match the results in the real world. Since even very simple starting hypotheses can lead to very complex behaviours, it is not always clear how to interpret the results. On the other hand, this complex behaviour does characterise real systems, and that is one of the reasons that CAS simulations fascinate so many people. To illustrate, I will discuss some typical examples of such simulations.

6.4 Cellular automata

A cellular automaton is a highly simplified representation of certain dynamical processes that can demonstrate unexpectedly complex behaviour. An agent here corresponds to a "cell" on a type of geometric grid of cells (in one, two or more dimensions). A cell only interacts with neighbouring cells. At every step or sequential point of time, every cell will automatically change state, depending on the states of the neighbouring cells. The state can be "active" or "non-active", but can also be a choice between a more complex series of options. Although simple and deterministic for each individual cell, the evolution of all cells together is in practice generally very complicated and unpredictable (chaotic), and exhibits a virtual limitless variety of behaviours.

This wealth of forms and behaviours has led the mathematician Stephen Wolfram to claim that all natural phenomena and the entire universe can be reduced to cellular automata. Although his (voluminous) book *A New Kind of Science* gives plenty of examples of cellular automaton models of natural phenomena, his statement has as yet not impressed the scientific world. A reason for this is that cellular automata are still too close to Newtonian reductionism, and ignore the more complex, asynchronous and indeterministic interactions that operate in the real world.





Example: "Game of life"

This is a two-dimensional "chess board" with square cells which can be active (black in the illustration to the left) or inactive (white). The rules are as follows:

- if a cell is inactive and three of its neighbours (of the eight neighbouring cells) are active, it will become active ("alive") in the next step
- a cell remains active if two or three neighbours are active
- in all other cases, the cell becomes (or remains) inactive.

If you allow the game to evolve from different initial configurations, you will see all sorts of patterns of activity, which stabilise, repeat themselves, advance, reproduce, "die out", and show other behaviours that remind us of living beings, but that are in fact very simple. Because of this endless variation, the game has long been popular as a screensaver. For an interactive demonstration, see: http://bitstorm.org/gameoflife/.

6.5 Swarms

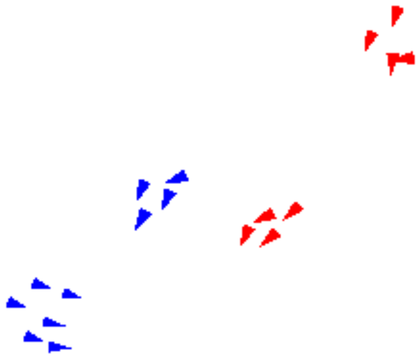
A swarm is a simple, but realistic, representation of the collective behaviour of animals that move in groups, such as flocks of birds, schools of fish, swarms of bees or herds of sheep. The agents (the individual animals) are represented as points that move in a two- (or three-) dimensional space. For a beautiful interactive demonstration, see: http://www.red3d.com/cwr/boids/applet/.

- behavioural rules for an agent in a herd:
  - the agent always tries to move towards the place where most of the other agents are
  - the agent aims to maintain minimal distance from other agents

Result: the agents form an irregular herd with those on the edge continuously trying to get to the centre, but in doing so, push other agents away. This is a realistic representation of the behaviour of large groups of herbivores (e.g. sheep, zebras, wildebeests) on an open plain, where they can be attacked by predators such as lions. Those on the edge of the herd are after all the most vulnerable to attack, while those in the centre run the least risk. On the other hand, they cannot stay too close together, because then there is not enough grass to eat for each individual.

- additional rules for agents in a swarm:
    - the agent tries to move in the same direction and with the same speed as the average movement of its neighbours
    - the agent maintains a minimal distance from obstacles

Result: the agents form a gracefully moving, beautifully synchronised swarm, that regularly changes direction and flies around obstacles by splitting into two sub-swarms that flow together again after the obstacle. This is a simplified representation of the collective behaviour of "flocking animals" that are in constant movement, such as flying birds or swimming fish.



- Possible additional behavioural rules with two types of agents:
    - the "predator" agents fly towards the closest prey
    - the "prey" agents flee from the predators

Result: the simulation produces breathtaking collective "hunting scenes". Think of a school of small fish, like herring, which is attacked by a group of sharks or tuna. The school of herring fans out when they perceive a tuna fish in their midst, but is corralled again when the tuna close in from several angles. During this, individual herring keep trying to be in the middle of as many other herring as possible, to reduce the chance of being eaten by a tuna.

- An unusual variation on the swarming rules

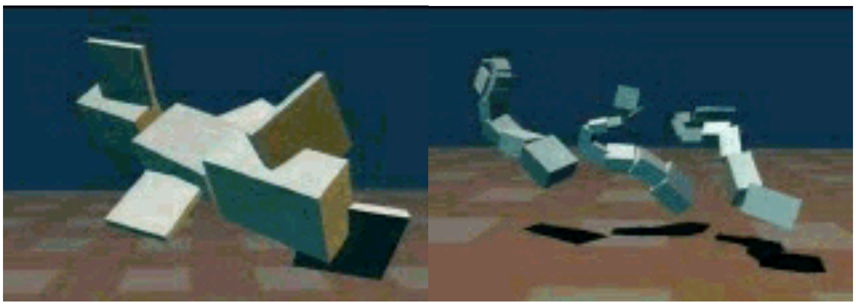
We assume again that agents are guided by the average speed and direction of their neighbours, and that they maintain minimal distance to avoid collisions, but now give them two different targets: the "red" agents aim to go up as much as possible, and the "blue" aim to go down. Red and blue agents are thus in each other's way and will be inclined to join their own kind that is moving in the right direction. However, when a red agent manages to reach the top layer, it will turn into a blue agent. In the same way a blue agent that reaches the bottom layer becomes a red agent.

If you have paid attention, you may recognise an earlier example: the emergence of convection currents in a fluid that is heated from below: the red agents represent the warm particles in the fluid, and the blue agents represent the cold particles. Although I have not yet seen such a simulation, I would expect, provided the rules have been set precisely (e.g. concerning speed and mutual distance), that the result would indeed be that the "swarm" self-organises in a regular, rotating pattern of Bénard rolls. This illustrates the universality of the concept of an agent, which is not restricted to the representation of living beings. It also shows that the thermodynamic approach of self-organisation and the more modern CAS approach are perfectly compatible.

6.6 Artificial life

The field of artificial life aims to understand the fundamental behaviours and evolution of living beings and ecosystems, using computer simulations (or sometimes small robots). There are many different ways in which this can be done.

A classic, but still very impressive, illustration can be found in the work of computer artist Karl Sims (see http://biota.org/ksims/). In a virtual, three-dimensional environment, he has allowed animals to evolve that need to perform certain physical activities, such as swimming, walking, or fighting for food, as efficiently as possible. The simulated environment is subject to realistic physical forces, such as gravity on land and viscosity (fluid resistance) under water. The bodies of the agents consist of a combination of square blocks of different sizes and shapes, which are stuck together with "joints", so that one can move relative to the other. The movements are controlled by an elementary "nervous system", which connects observed conditions with possible actions. The body and nervous system structures are initially random, but subsequently evolve under influence of variation and selection. In the simplest scenario, the selection criterion is the speed with which the creature manages to move, taking into account physical forces. On land, this produces creatures that walk with two or more "legs", some that shuffle forward like a snake, or even some who repeatedly allow themselves to tilt and "fall over", like an acrobat that forms a wheel by falling from the feet to the hands and again to the feet. Under water, this similarly leads to a variety of swimming styles, which usually resemble those of existing animals, but which can be very inventive—although they remain plausible. In the illustration below you can see some of the "swimming" creatures (a snapshot from the animated movies that you can download on the web). In a "combat situation", agents have to try to pull a piece of food towards them as fast as possible, before their competitor can get to it. Here, evolution again leads to different strategies, sometimes based purely on speed, sometimes based on subtle tactics to trick the competitor out of the piece.



Most artificial life simulations are however more abstract, and interested in general evolution and survival strategies in a simplified environment. This often takes the form of a two-dimensional grid, like with a cellular automaton: agents are spread across different cells, and can move by regularly "transiting" to a neighbouring cell, as long as they follow the condition-action rules. An example of this is the Sugarscape environment, where food ("sugar cubes") is spread out over a two-dimensional landscape, and where agents aim to obtain as much food as possible.

The most universal situation, first programmed by biologist Tom Ray in his famous *Tierra* simulation, is the following:

- Agents aim to collect "food", "energy" or "fuel". There is a limited amount of food in circulation, and agents can "steal" food from each other, or even "eat" each other, to recuperate food reserves of others.
- Agents that find insufficient food to survive, die, so that natural selection takes place and only the best adapted survive. (This selection is more "natural" than Sims', because it has not been concretely specified *what* the agents have to do (e.g. swimming or walking) to be selected.)
- Agents that find more than a minimum amount of food have an energy surplus and can use that to procreate.
- During procreation, small "mutations" are made in the behavioural rules according to which the descendants will try to get food.

Result: we see the evolution of a growing diversity of agents of various "kinds", which live for example as "plant", "herbivore", "parasite" or "carnivore" and thus compete or cooperate with each other. These species have each adapted in their own way to the ecosystem that they together form.

A fundamental question that needs to be addressed is to what extent agents and/or the ecosystem become more complex as the simulation lasts longer. In a sufficiently sophisticated environment (such as Tierra), the ecosystem and some of the agents do indeed become more complex, but this process eventually reaches a limit after which it appears that complexity no longer increases. Thus far, the virtual organisms are still much simpler than the simplest living being. This has caused artificial life researches to think more fundamentally about the evolution of complexity, hoping to come to an open-ended (unlimited) simulation of evolution. This should at least be able to produce consecutive hierarchical system levels (e.g. cells and multicellular organisms). We will discuss this problem later in a more abstract way, when we consider supersystem transitions.

6.7 Simulated societies

Multi-agent simulations have become popular recently in sciences such as economics, sociology and psychology, in order to understand social interactions and the formation of markets and societies. The starting point is similar to that of an artificial ecosystem, except that the agents now all belong to the same "species", and are therefore not supposed to "eat" or eliminate each other. Rather than striving for "food", agents strive for "benefit" or "reward" (which in the simulation mostly amounts to the same thing). The emphasis is on the strategies that agents use to do transactions with each other, and how from those individual strategies, collective or social organisation arises. Such strategies to obtain as much "profit" as possible from the interaction with others have already been described mathematically in game theory, developed in 1947 through a collaboration between the cybernetician and mathematician John von Neumann and the economist Oskar Morgenstern. But as soon as more agents take part who do not all follow the same rules, the situation becomes too complex to describe mathematically. In that case, we switch to simulations.

The starting point that has been used most often is the *prisoners' dilemma* game, with the following rules:

- an agent can help another agent (cooperate) or betray it (defect)

- if both help each other, both are rewarded
- if both betray each other, both are punished
- if one helps and the other betrays, the traitor gets the reward (which is larger than if he would not have betrayed), while the helper loses out (more than if he would not have helped)

The intention is that an agent accumulates as much benefit as possible, and as little disadvantage as possible. Different agents have different behavioural rules or strategies to play the game: depending on what the other agent did in the previous "moves" of the game (help or betray), the agent decides how to behave.

The problem with the Prisoners' Dilemma is that it is collectively best if you help each other, while individually, the best strategy is to betray and hope that the other will not do the same. The goal of the simulation is to find out if collective cooperation can arise even if individual profiteers have the largest advantage in the short term. This simulation was first set up by political scientist Robert Axelrod, as described in his book *The Evolution of Cooperation* (1984). Axelrod turned the game into a tournament, inviting different specialists in game theory to come up with strategies to gain as much profit as possible in consecutive game rounds (an agent was always confronted with different opponents).

<u>Result:</u> the agents that followed the very simple tit-for-tat strategy (if you help me, I help you, if you betray me, I betray you), did best on average (and thus won the tournament). The reason is that profiteers do not gain any advantage when they play against a tit-for-tat agent, because they get the same treatment, while tit-for-tat players who played against helpers or each other got a benefit out of their cooperation.

To discover whether evolution also leads to more cooperation in the long run, we introduce some more complex variations: those agents that gain a lot of advantage reproduce, the others die out.
<u>Result</u>: in the end only tit-for-tat is left and everybody helps everybody.

If we allow mutation of behavioural rules during reproduction, we also get agents that help in all circumstances. These will after all do equally well in a tit-for-tat situation, since everybody helps each other anyway. The presence of these "too good" agents does however create an opportunity for the evolution of deceivers who can exploit them without limitation. Consequently, the deceivers only need to compete with tit for tat, which suppresses their chances of winning, because tit-for-tat collect additional benefit when they play each other.
<u>Result</u>: tit for tat remains in the majority, but with fluctuating subpopulations of pure helpers and pure deceivers.

This looks like a simple but realistic representation of our society: most people are willing to cooperate, but expect something in return sooner or later (mutual altruism). If this expectation is not realised, their cooperativeness stops. Conversely, some people are naive or good enough to help in all circumstances. This enables a small number of deceivers to gain an advantage without doing something in return. It is these cheaters that remind the majority that they have to remain careful, even if they assume that most people are honest.

Let us look at more complex variations. If the agents only interact with their immediate neighbours, zones arise where everyone helps everyone, and other zones in which everyone deceives everyone, with transition zones between them where both cooperation and abuse of trust take place. This is not an unrealistic representation of today's world, in which most countries have a social, economic and legal system that guarantees reliable, cooperative transactions, but in which certain areas, such as Somalia or Afghanistan, live in an anarchy where only the right of the strongest counts.

If the agents are divided into different "groups" or "families" (recognised by specific labels), they will typically trust and help members of their own group, even if they have not yet had the chance to observe their behaviour, while they will *a priori* distrust members of other groups. The reason is that members of the same group will normally follow the same rules, so one can trust that they will respond to cooperative behaviour in the same way. The distrust of other groups whose rules you don not know, is a precaution that means that groups with deceivers stand no chance and will eventually be selected against, while groups consisting of only helpers will accumulate advantage and thus grow. This may explain why "failed states" such as Somalia and Afghanistan are a very small minority, partly because individuals who get the chance emigrate from those regions to areas that are more cooperative.

6.8 Simulated culture

Another interesting, but as yet less well-known application of the computer simulation of agent systems is the development of culture. Culture here can be defined as knowledge or behaviours that are transmitted from agent to agent. As far as I know, the first of these simulations, *Meme and Variations*, was programmed by Liane Gabora, a Canadian researcher who is affiliated with the Centre Leo Apostel of the Free University Brussels. The fundamental rules of this and similar simulations are as follows:

- agents look for the best solutions to problems
- they can try to find a solution themselves through trial and error
- they can also copy the solution of the best of their neighbouring agents by imitating it

Result: the group finds the best solutions if they partly imitate others and partly invent individually. If they only imitate, there is no creativity and the existing solution cannot be improved. If they only try individually, a lot of energy is wasted on failed attempts and unnecessary repetition, as they each have to reinvent the wheel. In the ideal case, achieved by experimenting until the optimal proportion of imitation to innovation has been found, good solutions will quickly spread among the "population", but without precluding other agents from finding a better one.

In a more complex variation, agents will not slavishly imitate their neighbours, but assess different ideas from different agents, based on certain criteria. They will then adopt the most "convincing" ideas, and in turn try to transmit these to others. This way, the most convincing ideas will quickly spread among the group, but enter into competition when one idea contradicts another. If we also allow variation of the spreading ideas (memes), a simulation of a cultural evolution appears, such as can be seen with the diffusion of fashion phenomena, jokes, or rumours.

My colleague at the Free University Brussels, social psychologist Frank van Overwalle, has developed, with my input, a psychologically realistic simulation model for this, called *Talking Nets*. It attaches special importance to mutual trust among agents. The fundamental idea is that you will give little credence to the utterances of someone that you have often disagreed with in the past, while you will listen carefully to the ideas of someone that you normally agree with. On the other hand, someone that you normally disagree with will, knowing your scepticism, tend to make extra effort to convince you after all. The result is a complex dynamic, which can lead to the division of a group into "subcultures" which have very different opinions.

Chapter 7. Complexity: fundamental concepts

7.1 A relational ontology

Now that we have discussed the historical developments of different approaches to the problem of complexity and evolution, it is time—using the insights that we have gained—to approach the foundations of the issue. This will allow us to formulate the core concepts that can be found explicitly or implicitly in these different approaches. Building on these foundations, we can then gradually try to work towards a coherent approach of the most advanced problems, such as the development of life, society or human intelligence.

Ontology can be defined as the branch of philosophy that studies the most fundamental categories of phenomena, that is, the essence or *being*. In other words, we search for the core components or elements of reality. In the Newtonian worldview, these are the *particles*, that is, small pieces of matter, and the energy fields or *forces* that cause the particles to move according to specific trajectories in space and time. So the Newtonian ontology is *materialistic*.

In the evolutionary-systemic worldview, this starting point is useless. After all, we want be able to describe *emergent* properties and *organisation*, independent of its material components. Therefore, we need to search for elements that are more abstract. These abstract elements must allow for complexity from the beginning, rather than immediately reducing it to independent building blocks. Still, the elements themselves have to be as simple as possible, because otherwise we risk drowning in complexity from the start.

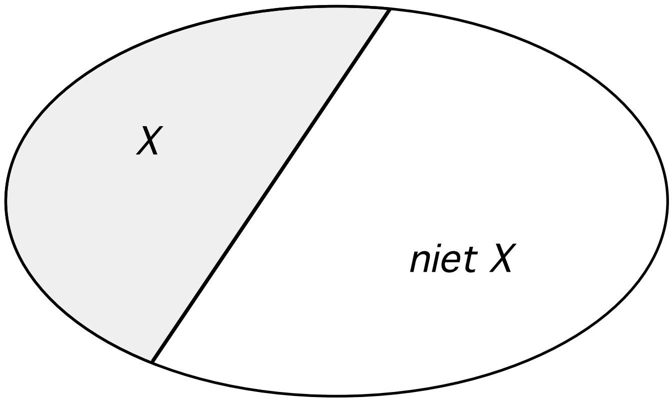
This can be resolved by making the elements *relational*, that is, intrinsically interdependent or connected, rather than independent. Every element can only exist *in relation to* another element, and the network of relationships that develops in this way is the basis for complex organisations.

7.2 Distinctions

Probably our most fundamental action is to distinguish. A distinction divides the world into two parts: that which belongs to the distinct category, and that which does not. Some examples: big—small, light—dark, left—right, alive—dead, beautiful—ugly, ... Distinct phenomena are considered *different* in at least one meaningful aspect. Phenomena that are not distinguished are implicitly considered *similar*, that is, belonging to the same category, with the same properties. A distinction is not a "thing", not an independent element (such as an elementary particle), but a relationship, namely the relationship of difference between a category of phenomena with a certain property and the complementary category of phenomena that lack that property.

If we would not make distinctions, it would be impossible to recognise any structure in, or say anything meaningful about, our perceptions or experiences. In other words, without distinctions every form of perception, thought or communication becomes impossible. We can only assign meaning to a phenomenon by placing it in a certain category X (that is, assign it the property X) and in that way show how it differs from other phenomena that do not belong to X. A distinction is therefore the most primitive element of every description or theory.

Note that a distinction does not need to be exact or absolute: intermediate cases are possible where we cannot say with certainty whether something does or does not belong to category X, but at best that it has the property X to a certain extent. For example, is a man of 1.75 meter “tall” or “not tall” (small)? At best, we can say that he is not very small, although we would not really want to call him “tall”. Conversely, we would classify a man of 2.20 meter as “tall” with no hesitation. Such categories are called *fuzzy* in mathematics: the dividing line between X and not-X is blurred, and where we place it will often depend on context.



The abstract notion of distinction can be found in all general representations, such as language, logic, mathematics or systems theory, as a foundation on which more concrete descriptions can be built. Some examples of such fundamental, abstract distinctions:

language:	yes	↔	no
knowledge:	true	↔	false
computing:	1	↔	0
logic:	proposition	↔	negation
mathematics:	set	↔	complement
topology:	inside	↔	outside
systems theory:	system	↔	environment
perception:	figure	↔	background

Although these classical representations, which are among others the foundation of the Newtonian worldview, explicitly assume distinctions, they generally forget to mention that distinctions cannot exist independently of one another. The Newtonian worldview is after all analytic or reductionist: it explains complex phenomena by dividing them into different elements or properties. Next to this reductionist model, we also need a holistic approach, which reconnects the distinct elements into a whole. This brings us to the second type of relational elements in our ontology: connections.

7.3 Connections

Distinctions are only meaningful if they lead to other distinctions. That is, a link or **connection** needs to exist between distinctions or categories, so that our knowledge about the one category can also tell us something about the other. These connections can be best expressed in the form of the condition-condition rules that we have introduced above, where a condition expresses that a perception belongs to a certain category. For example:

- IF something belongs to category X, THEN it also belongs to category Y: $X \rightarrow Y$
- IF something is a banana, THEN it is curved: banana → curved
- IF you let go of a heavy object, THEN it will fall: let go → fall

Such connections allow us to connect one category with another and to use our knowledge about the one to derive something we did not yet know about the other. Connections thus represent *interdependence* between distinctions.

Note that connections can be fuzzy, just as distinctions. In that case, given the first category (IF) we can say something about the second category (THEN), but we cannot do so with certainty. Example: banana → yellow. Not all bananas are yellow (some are green or brown), but given that an object is a banana, there is a higher than average possibility that the object will be yellow. As we will see later, such conditional probability means that the connection gives us **information** about the object, even if we have no certainty.

A distinction that does not lead to another distinction is worthless, useless, or meaningless, because it does not relate to the rest of the world. From a cybernetic point of view, we could say that the function of perception or communication is to give us potentially useful information, with which we could for example decide which action to undertake, or to predict which other condition will now also appear. In the words of cyberneticist Bateson: *a difference has to make a difference* to provide us with any information.

On the ontological level, this idea has been formulated most explicitly by the philosopher Leibniz as the principle of the **identity of the indiscernibles**. The principle states that if two phenomena, a and b, are in no way distinguishable or discernible, then they are identical ($a = b$), and there is actually only one phenomenon. Our original theory about two phenomena can thus be simplified by keeping only one. This is a very important epistemological *and* ontological principle, which is however often neglected, even by very intelligent thinkers.

<u>Example:</u> the "zombie"

This example has been devised by the philosopher Chalmers to try to show how difficult it is to understand consciousness. Chalmers defines a zombie as a being that looks exactly the same as a human being, and that behaves like a human in all possible ways, but that has no feelings or "consciousness" while performing its actions—a bit like a robot that mechanically carries out tasks according to its programming, without any emotions. "What is it that a real human has that a zombie lacks?", Chalmers then wonders. He defines this as "the hard problem of consciousness", which by definition cannot be solved in a scientific manner, since science only works with observable distinctions.

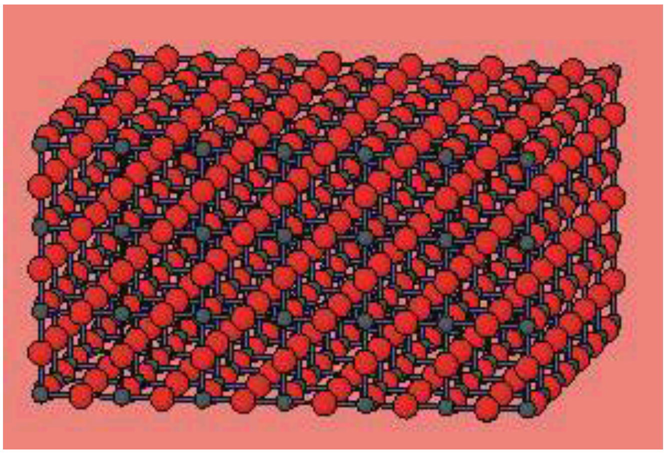
If we apply the principle of Leibniz, however, normal humans and zombies are identical, since it is impossible to distinguish the one from the other. The distinction "consciousness ↔ no consciousness" is only meaningful if beings without consciousness can be distinguished in some way from others, such as in the way in which they behave. This is for example how we determine whether someone is conscious, or if someone is aware of a situation, such as a fast-approaching car. If we would not do that, we might as well assume that all people other than us are zombies, since we can never know for certain whether they really have the same feelings as we do, or whether they only simulate them. Chalmers' "hard problem of consciousness" is thus, according to Leibniz's principle, a false problem, that needlessly complicates matters.

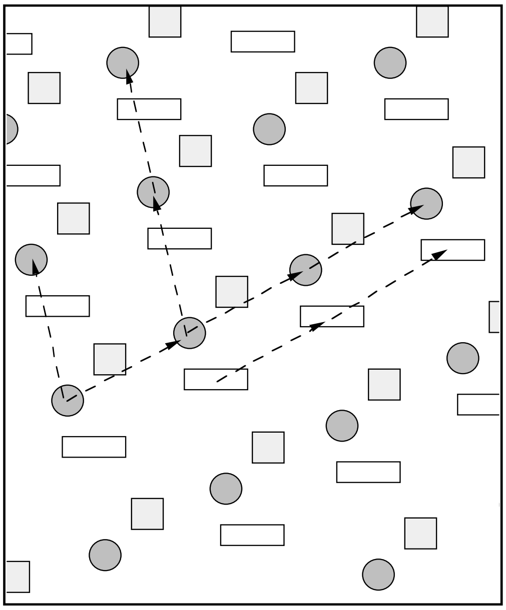
Although it may not be evident for everybody that distinctions are only meaningful if they are connected, the converse is self-evident: connections are only meaningful if there are different phenomena to connect. We can therefore decide that distinctions and connections are both necessary to describe reality.

7.4 Order

One concept that always emerges in discussions about complexity, self-organisation and related phenomena is the concept of **order**. In practice, however, it turns out that nobody defines this concept very well. Intuitively, we associate order with *regularity*, i.e. the following of set rules. For example, the magnets in the magnetised material (section 3.2) all point nicely in the same direction and none deviate. In a crystal (another example of self-organisation), all molecules are at the exact same distances from each other, in a perfect geometric pattern (see illustration below).

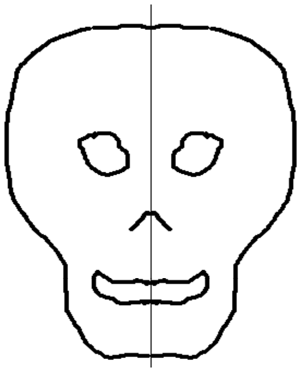
Mathematics offers a precisely defined concept to describe these kinds of situations, namely *symmetry*. A pattern is symmetrical if it is invariant under certain transformations. A transformation is a mathematical function that maps one part of a pattern onto another part. Invariance means that such an internal shift does not alter anything in the pattern. This is only possible when the different parts of the pattern are identical to each other.





This is best illustrated using an example. In the illustration to the left you see a regular pattern that one could find for example on a piece of wallpaper. The pattern is symmetrical, because we can move it in the direction of the arrows, without changing the way it looks. That is because the circles, squares and other elements of the pattern are repeated at equal distances, in the direction of the arrows. This is the same kind of "translational symmetry" that you find on wallpaper or in crystals.

Another example is mirror symmetry, such as that in the drawing of the skull below, where the left half is exactly the same as the right half, except for the mirroring. Symmetry means that to reconstruct the whole pattern, it suffices to know a part of it. For example, the left half of the skull suffices to reconstruct the right half, while for the "wallpaper" a part of the pattern with one circle, square and rectangle suffices to generate the rest of the pattern through shifts over regular distances.



One part of the pattern determines the configuration of the rest, barring a small "shift". This is a special case of **connection**: one component determines the other. IF we know the configuration of the first pattern, THEN we also know all the others. The repeating patterns are all connected, but intrinsically they hardly differ.

If we now generalise this property to order overall, we come to the following conclusion: *order is characterised by many connections, but few distinctions*. In an ordered system, the components are homogenous or repetitive. The prototype of such a system is a *crystal*: here, all molecules are identical, oriented in the same direction, and positioned at equal distances from each other. If you know the configuration of one molecule, you know all of them.

Now extend this definition of order to the *limit of maximal order*, that is, only connections, no distinctions. This means that every part is the same as every other part, no matter how small the observed part is. In other words, there is no distinction or differentiation at all; everything is perfectly homogenous, you cannot distinguish anything from its surroundings. This equals perfect *emptiness* or vacuum. It follows from the principle of the identity of the indiscernibles: if no components or parts can be discerned, there are no parts, and therefore there is nothing.

7.5 Disorder

Disorder, another core concept in complexity science, can now be defined as the opposite of order. This means that *disorder is characterised by many distinctions, but few connections*. In a disordered system, the parts are different and independent from each other. The prototype of such a system is a *gas*: the gas molecules move completely independently from each other and all have different speeds, directions, positions and mutual distances. Even if you know the state of one gas molecule perfectly, you still do not know anything about the other molecules.

The peculiar thing however is that a gas is *statistically* homogenous: the *probability* of finding a certain molecule in a certain place is the same for all places. If this were not the case, by definition there would be a statistical dependence stating that one place differs in some systematic way from another place, and therefore a connection between these places. The law of large numbers states that if we consider a very large number of cases, then what is probable becomes what happens in practice: because there are so many molecules, the average (expected) number of molecules in a certain place is in practice virtually equal to the actual number of molecules. Since the probabilities and therefore the averages are divided homogenously, this also means that the molecules are divided virtually homogenously. This is why a gas has no structure or form.

Now let us discuss the *limit of maximal disorder*: all distinctions, no connections. This means that every part, no matter how small, is different and independent of every other part. It is difficult to imagine such a situation. Yet this is a good description of the vacuum as it is postulated in quantum mechanics. According to this theory, quantum fluctuations constantly produce "virtual particles" in every point of empty space. These almost immediately disappear again, without leaving any trace. Every point in space is thus in a constant flux, independently of any other point. Yet in practice, such a space appears about as empty as the empty space of Newtonian mechanics.

A more concrete illustration is the pattern of "snow" that you can sometimes see on a television screen when it does not receive a signal. In this case, the television tries to create pictures based on the "noise" of electromagnetic waves in the background. Each pixel is independent of every other one, and the colour value of this pixel changes continuously and unpredictably, creating the impression of whirling dark and light spots. However, if you look at the screen from further away, so that you can no longer distinguish the individual pixels, it appears as a uniform grey. The accompanying sound is called white noise. When this white noise is not too loud, your brain will soon stop paying attention to it, making it appears as if there is no sound at all.

Conclusion: in the limit, we again end up with emptiness. This means that the limits of maximal order and maximal disorder become one. It is only when order or disorder are imperfect that we can distinguish them.

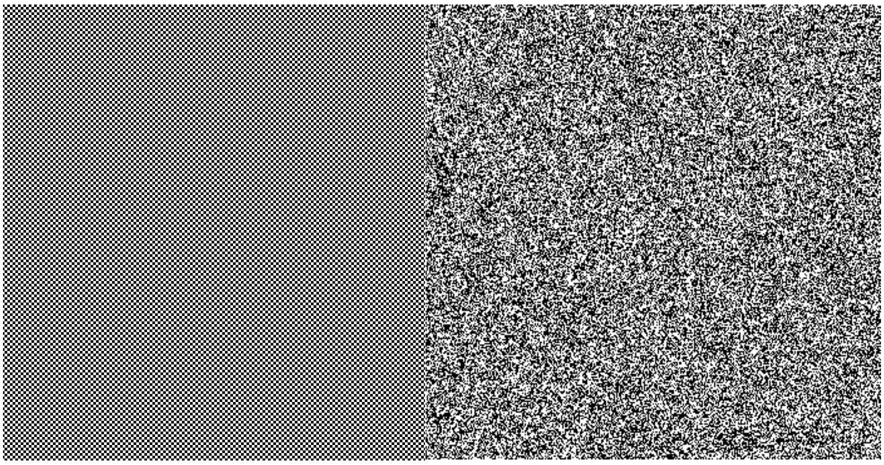
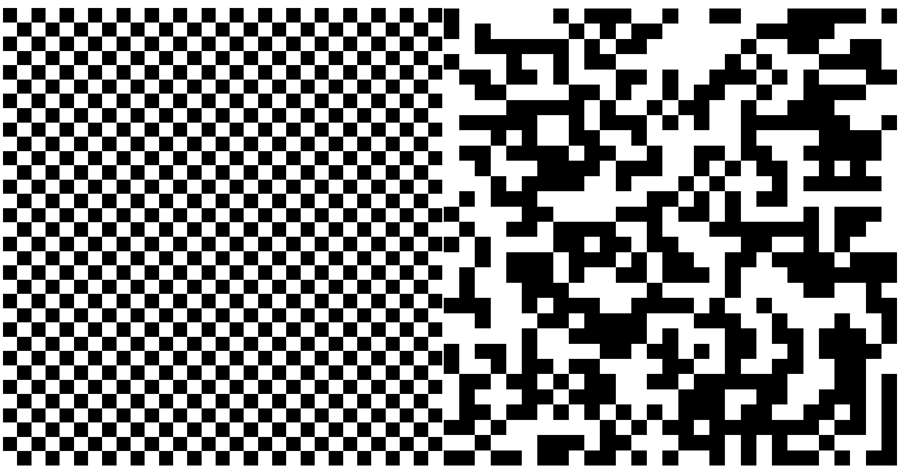


Illustration: order (left) and disorder (right) both produce a uniform grey when they are seen from a large enough distance.

However, if you zoom in (illustration below), and look at both patterns from close by, it becomes clear that the pattern on the left is perfectly predictable or regular, and the pattern on the right is completely unpredictable.



7.6 Complexity

- What is **complexity**?

The word "complex" is derived from the Latin *complexus*, which means folded together, entangled or embracing. This means that something is complex or "complicated" if it consists of two or more parts that are connected to, or entangled with, each other in such a way that it is hard to take them apart; distinct components are stuck together. A complex is thus characterised both by **distinctions** and by **connections**. Later, "complex" also got the meaning of "difficult", since it is difficult to analyse systems or problems with many mutually dependent components, and thus difficult to understand or solve them.

To this, we can add that something becomes more complex the more distinctions and connections it has. Based on our earlier analysis, we can then decide that complexity exists *in between order* (many connections) *and disorder* (few connections). In the CAS approach, this intermediate zone is also called *the edge of chaos*, meaning the wilderness between rigid, "frozen" order and turbulent, chaotic disorder. This is the area where "interesting" phenomena happen: self-organisation, life, evolution, thought... Neither a crystal, where everything is perfectly ordered, nor a gas, where everything whirls together uncontrollably, is really interesting.

This is not just a subjective assessment. As can be seen from the illustrations of order and disorder, our brains are set up in such a way that they soon stop paying attention to uniform patterns, whether these are intrinsically orderly or disorderly. A perfectly clear tone (sinusoid) or a very regular hum (such as the buzzing of a refrigerator) will soon become unnoticeable, just as much as a constant murmur. These kinds of background noises will typically only become noticeable when they stop suddenly, that is, when the homogeneity is broken. Likewise, your eye will pay no attention to a uniform, unchanging background, but direct itself to that which breaks the symmetry, that which stands out as being different or unexpected. The more complex the structure, the more it attracts attention—either consciously or subconsciously.

- Examples of complex systems:

- The tropical rainforest: millions of different plants and animals (distinctions), that are all in some way or other dependent on each other (connections).
- The brain: billions of neurons (distinctions), connected by trillions of synapses (connections).
- Society: billions of individuals that are all doing something different (distinctions), but still influence each other in wide-ranging ways.

- How do we measure complexity?

There have been dozens of attempts to define complexity in a measurable, quantifiable manner, in order to determine the exact complexity of a phenomenon. None of these definitions is however universally applicable.

One of the most well known definitions is *the length of the shortest possible description* of the phenomenon. The idea is that the more complicated a phenomenon is, the harder it will be to describe it completely. The problem with this definition lies in the language used for the description. Certain languages (such as mathematics) can describe certain phenomena efficiently (e.g. a complicated geometrical figure such as a fractal), but are very inefficient for others (e.g. romantic feelings). For any given description, you will never be sure that this description cannot be shortened by using another, more efficient, language.

The reason for these difficulties is that complexity is not really measurable. There are after all many different ways to add distinctions and/or connections and thus increase complexity, but these cannot be compared with each other. Trying to add the total number of distinctions and connections is a bit like adding apples and oranges. For example, is a helicopter more complex than a plane or a tank? These phenomena are not really comparable. A motorbike, on the other hand, is more complex than a bicycle, since it contains all the parts of a bike, with a lot of extra parts besides. We may decide that complexity can at most determine a *partial* order relationship. This means that not everything can be ordered according to degree of complexity: *A* can be more complex than *B*, or vice versa, but *A* and *B* can also be incomparable.

7.7 Differentiation and integration

We are however not so much interested in the static complexity of a phenomenon, but in its evolution. We especially want to ascertain whether complexity increases rather than decreases. According to our definition, complexity increases if the number of distinctions and/or connections increases.

Increase of distinctions with decrease of connections means increase of disorder, not of complexity. An example is the melting of ice, where the water molecules, which initially were stuck together in the ice crystal, now begin to move independently. Another example is the disintegration of a society and its army into a number of competing fractions such as has happened in for example Somalia. Where the different groups were initially connected to each other through social rules and conventions, they now work each for themselves, which makes the situation in the country chaotic and unpredictable.

Increase of connections with decrease of distinctions on the other hand, implies increase of order. An example is magnetisation, where independently oriented molecules now all align themselves in the same direction. A social example is the development of a totalitarian system that prohibits deviant opinions, independent parties and private initiatives, and that subjects everyone to the same rules. This makes the society more stable and more predictable, but also leads to stagnation, because innovation is prevented.

Increase of distinctions is called **differentiation**, that is, increase of the internal diversity so that more different kinds of parts can be distinguished. Increase of connections is called **integration**, that is, the parts become more dependent on each other, and form a more coherent whole. **Complexification** (becoming more complex) can then be defined as:

differentiation + integration

A physical example is the Bénard self-organisation: the initially independent fluid molecules begin to coordinate their movements into a coherent flow (integration). On the other hand, the homogenous fluid divides into independent cells or rolls, each with its own flow pattern (differentiation). A biological example is the development of an embryo from a homogenous clump of cells to a baby that is ready to be born: the cells begin to divide into different types (e.g. skin cells, nerve cells, bone cells, ...) and thus produce organs that will form a coordinated, coherent whole. In history, we see such complexification in the evolution from groups of hunter-gatherers to agrarian and later industrial societies. The initially interchangeable individuals begin to perform more and more differentiated, specialised functions (e.g. warrior, priest, merchant, solicitor, architect, ...), but become more dependent on each other and on social structures (solicitors can for example not produce their own food, but need to rely on a network of farmers, transporters, shops, etc.).

7.8 Organisation

The last term that is frequently used in the context of complex systems is **organisation**. We can define this as:

> *structure with function*

An example is a human organisation, which consists of individuals who interact according to certain communication channels (structure), and who do this in order to achieve a certain economical or social goal (function). A structure consists of a number of distinguishable parts that are connected to each other. "Structure" implies therefore a minimal degree of complexity, i.e. both distinctions and connections, but without emphasising the *number* of distinctions or connections. "Function" means that this structure has a certain goal or utility.

Examples:

- The brain clearly has goals; the rainforest does not.
- A machine (such as for example a watch) has a function; a complicated assembly of loose parts does not.

Both the brain and the watch show organisation. Although a rainforest is far more complex than a watch, it is still not organised. Note that according to this definition, the traditional examples of self-organisation, such as crystallisation, magnetisation, or Bénard convection, can better be placed under the denominator of "self-ordering" or "self-complexification", since the created structures do not have a purpose.

Note: The most realistic complex systems do however at least have an implicit purpose: to survive. Without this property, natural selection would have eliminated them a long time ago. The different parts and connections of such a system can then be interpreted as contributing to that purpose. In that sense, the system does not only show complexity, but also organisation. In this broader sense, even the rainforest or the Bénard rolls can be seen as "organisations". The concept "function" with its connected concept "organisation" is therefore to a certain extent fuzzy, and depends on how we define "purposeful".

Chapter 8. State spaces

8.1 Models

I now intend to use the intuitive concepts of “distinction” and “connection” as basic concepts for the creation of a scientific model of an arbitrary system. A model is a simplified representation of a system, created by an agent, who is in this case the scientific observer. A concrete example of a model is a doll that represents a baby, or a map that represents a city. According to the scientific method, such a model is in principle meant to describe the system precisely and unambiguously, preferably quantitatively, although this description will in practice always remain an approximation.

For complex systems in particular, this approximation will only provide us with a very rough, almost caricatured sketch of what the system has to offer. A street map of a city tells us absolutely nothing about the people that live in this city or about the city’s climate, while it reduces its cultural life to the locations of the main museums and concert halls. This simplification of reality is an essential feature of a model: if the model would be as complex as the system itself, it would not give us any advantage and be unusable in practice. Imagine a map of a city on a 1/1 scale, with every trashcan, stone or weed depicted in all details!

This simplification means that different observers who are interested in different aspects of the system will create different models that are hard to compare. Maps made by geologists will note relief, water basins and soil composition, but will not list street names or the functions of buildings. The maps of sociologists will note the income categories or ethnic composition of the population in the different districts of the city.

The fact that a model is intrinsically limited and subjective does however not mean that it is scientifically unreliable or useless. The most important function of a model is to make predictions about how the system will behave in different circumstances, and in that way to help solve problems or achieve goals. A street map will for example allow you to predict that if you take the second street left and then the first right, you will be in front of the opera house that you are looking for.

In this book, it is not my intention to make such predictions, but only to illustrate the core concepts of a model, using some very simple examples. This will enable us to understand and define a number of fundamental concepts. It is also not my intention to go into technical detail or make complicated calculations. It will suffice to show that this is in principle possible, for those who want to accurately describe, predict, design or simulate a system. Thus we will get a foundation that allows us to better understand existing scientific models.

8.2 Objects

Every model of a system starts with one or more objects. Objects are the primitive components of a system model, the stable “things” or elements in which one is interested. An **object** is a distinction of the type “system—environment”, in which the object of investigation (the system) is distinguished from everything that does not belong to the study domain (the environment). An object is a system, but without consideration of its subsystems. The first simplification of the model consists in seeing the objects as elementary, i.e. without distinct parts. Some examples of possible objects, depending on the type of model, are: “building”, “person”, “particle”, “car”.

To belong to the type “object”, a distinction has to be *stable*. Objects remain equal to themselves. They are invariant. This means that they do not change under influence of all sorts of transformations or manipulations of the system. Although the properties of an object can change, we assume by default that the object itself remains. For example, a car can change place, drive faster or slower, or even be repainted in a different colour, but it remains the same car.

<u>Example:</u> the billiard ball

Billiard balls that move across a billiard table form a very simple example that we will discuss in detail to clarify the properties of a model. An individual billiard ball is clearly an object: the distinction between inside (“system”) and outside (“environment”) remains, whatever the position or movement of the ball. I can manipulate the ball in many ways, but the ball itself remains.

Note that the ball could in principle break into pieces, melt or evaporate (i.e. split up into its molecular components). This is however very unlikely, and a typical model of a game of billiards will not consider these possibilities. Objects are—like all elements of a model—idealised, simplified representations of an infinitely complex reality.

In a complex system, you can generally distinguish *different* objects or elementary subsystems. For example, the different billiard balls on a table together form a system. Elementary systems, such as particles, do not have subsystems, and form objects by themselves.

8.3 Properties

Properties can be defined as distinctions, attributed to one or more objects, that vary across different situations. These are also called *degrees of freedom*, because the object has the “freedom” to assume different values of a property. For example, someone’s weight (property) can vary from week to week, but the person (object) remains the same. The position (property) of a car changes with every movement, but the object “car” remains invariant. Let us consider the primary types of properties.

- Binary properties

These are properties that the object at a certain time either has or does not have. The property therefore has two possible values: *present* or *absent*. In logic, a binary property corresponds to the elementary predicate P that is attributed to an object (sometimes called subject) o: $P(o)$.

<u>Examples</u> of binary properties: “red”, “heavy”, “high”, “beautiful”...

- “Attributes”

These are properties with more than two possible values.

<u>Example:</u> temperature with values “freezing”, “cold”, “tepid”, “warm” and “hot”, or values ... 5ºC, 6ºC, 7ºC, ...

The number of possible values of a property can be finite or infinite. An infinite number of values can be *countable* (discrete) or *continuous*. For example, the natural numbers, 1, 2, 3, 4, ... are uncountably infinite: there are gaps between the consecutive numbers, so that we can list or count them one by one. Real numbers, on the other hand, do not have gaps between consecutive values: they follow each other uninterruptedly and can therefore not be counted.

Example: the position of a billiard ball on a billiard table in the longitudinal direction has in principle a continuous infinite number of values.

- Relationships

Relationships are predicates with more than one object: *P*(*a*, *b*, ...). Typical relationships have two objects, between which they establish a link.

Examples:

- *Bite* (*man*, *dog*)—this means: the man (object) bites (relationship) the dog (other object).
- *Left* (*a*, *b*)—ball *a* is located to the left of ball *b*.
- Ball *a* moves faster than ball *b*; ball *a* is located at a distance *x* of ball *b*.

8.4 States

At a certain point in time, an object *a* may or may not be characterised by a certain property *E*. This determines the elementary expression or *proposition*:

E(*a*) : "*a* has the property *E*"

This proposition is *true* if *a* does indeed have property *E*, otherwise it is *false*.

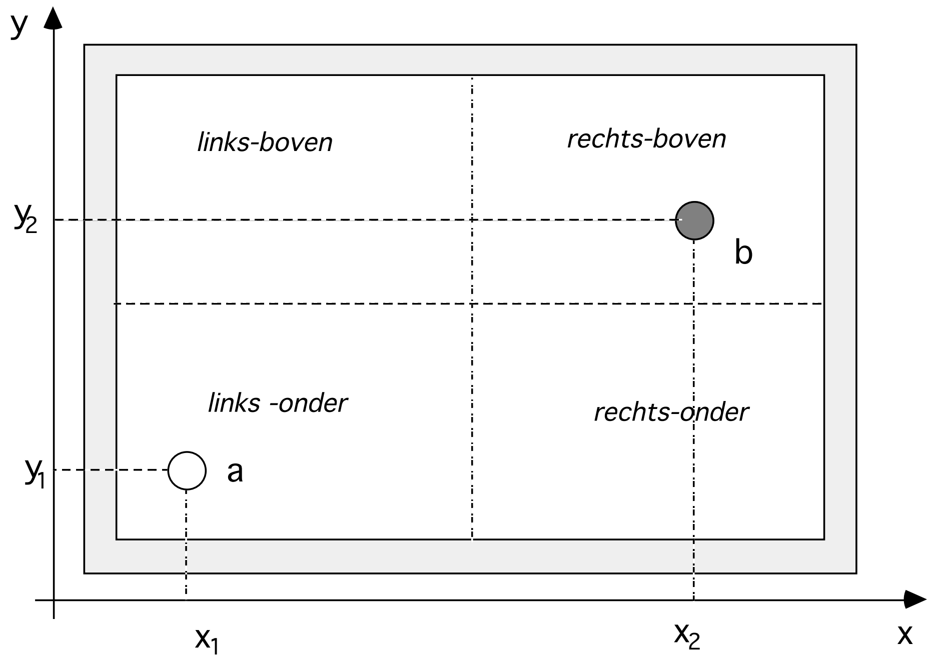
At any point in time, a system can be described using a number of such elementary expressions, that each ascribe properties to objects that are part of the system.

If there is only one object, one can leave out said object, because in that case propositions correspond unambiguously with properties.

Example: the proposition "ball *a* is located at the bottom left" can be shortened to "bottom left" if there is only one ball and therefore no ambiguity as to which object is positioned at the bottom left.

Elementary propositions can be combined into compound propositions with the help of logical operators: conjunction ("and"), disjunction ("or"), negation ("not"). The conjunction of all propositions that are true for the system at a given point in time defines the **state** of the system. For example, "ball *a* is located in position (x_1, y_1)" *and* "ball *b* is located in position (x_2, y_2)."

The state of a system gives us all *relevant information* about the system at that moment. "Relevant" here means necessary to solve problems or make predictions regarding the system. The relevance depends on the model and the specific problems it addresses. For example, the speed of a ball is relevant if we want to know where the ball will be located in the next moment; but the colour of the ball is irrelevant in that situation. In a model of a billiard ball that is only interested in the physical movements of the balls, the state will normally not include the property "colour". However, if we are also interested in the rules of the game, and in which strikes are performed by which players, colour becomes relevant and therefore will be part of the model.



The definition of a state thus depends on which problems you want to solve. If the state provides insufficient information to answer the posed questions, you will generally need to distinguish additional properties (or possibly objects), and add these to the model. For example, the position of a ball alone is insufficient to predict movements: you will also need the speed (or momentum) of the balls. Position and momentum together determine the state of a billiard ball in this case. In the Newtonian worldview, it is assumed that you are able to make *all* relevant distinctions so that the model becomes deterministic and can answer all questions. Later theories, such as quantum mechanics and chaos theory, have however shown that you can never hope to gain complete information about a system, and that a model will always remain an incomplete representation of the behaviour of a system.

8.5 State space

The **state space** of a system is the *set of all possible states in which the system can find itself.* This is a generalisation of our intuitive concept of the concrete, three-dimensional space in which we can move around freely to the abstract set of states between which a system can "move" when its properties vary. State space is sometimes also called *phase space* or *configuration space*.

The state space is generally indicated with the uppercase letter S; the individual states with lowercase letters: s_1, s_2, s_3, ...

$$S = \{s_1, s_2, s_3, ...\}$$

Suppose that our model contains N elementary propositions. Then there are 2^N (2 raised to the power of N) possible "true-false" combinations, and therefore 2^N states.

Example: one object with three binary properties: "left–right", "above–below", and "in front–behind". That means there are $2^3 = 2 \times 2 \times 2 = 8$ possible combinations of the different values of these properties and therefore eight states:

1)	left-	above-	in front
2)	left-	above-	behind
3)	left-	below-	in front
4)	left-	below-	behind
5)	right-	above-	in front
6)	right-	above-	behind
7)	right-	below-	in front
8)	right-	below-	behind

More generally: consider one object with different properties, each with a number of values n_1, n_2, n_3 ..., , then the number of states is equal to the product of the number of possible values for each property:

$$n_1 \times n_2 \times n_3 \times ...$$

- System consisting of several objects

Consider several objects a, b, c, ..., each having one property with a number of possible values n_a, n_b, ... Again the number of states is equal to the product of the possible values:

$$n_a \times n_b \times n_c \times ...$$

For several objects a, b, c, ..., each with several properties n_{a1}, n_{a2}, n_{a3}, ..., n_{b1}, n_{b2}, n_{b3}, ..., the number of states is again the product of the number of values for all objects and all properties:

$$n_{a1} \times n_{a2} \times n_{a3} \times ... \times n_{b1} \times n_{b2} \times n_{b3} \times ...$$

Each property of each object determines a dimension of the space. The number of dimensions (dimensionality) of the state space is equal to the number of properties or degrees of freedom. For example, a billiard ball that can move from left to right (along the x-axis) and from top to bottom on a table (along the y axis) thus has two independent properties (ignoring speed for the moment) and thus moves in a two-dimensional state space. A system consisting of *two* such balls however has $2 \times 2 = 4$ dimensions or degrees of freedom. The more objects a system has (or the more properties an object has), the more dimensions the state space has.

Mathematically, we can formulate this as follows:

The state space S of the composite system is the *Cartesian product* of the state spaces S_a, S_b, ... of the individual objects:

$$S = S_a \times S_b \times S_c \times ...$$

An individual state s is then represented as:

$$s \in S = (s_a, s_b, s_c, ...) \text{ with } s_a \in S_a, s_b \in S_b, \text{ etc.}$$

Reminder: the Cartesian product of two sets $A = \{a_1, a_2, a_3, ...\}$ and $B = \{ b_1, b_2, ...\}$ is equal to the set of all possible pairs formed from a combination of an element from the first set with an element from the second set:

$$A \times B = \{(a_1, b_1), (a_1, b_2), (a_1, b_3), ..., (a_2, b_1), (a_2, b_2), ...\}$$

A state of an arbitrary system can therefore always be written as a list ("vector"), consisting of the consecutive values of each of the properties for each of its objects.

$$s = (s_{a1}, s_{a2}, s_{a3}, ..., s_{b1}, s_{b2}, s_{b3}, ...)$$

8.6 Distance metric*

A state space is more than a set of separate, independent states: states can be located closer to or further away from each other. There is a distinction between states that are near ("in the neighbourhood") and those that are far. The existence of such distinction between near and far defines a topological structure. This transforms the set into an elementary "space".

It is possible to define an elementary measure (in mathematics called "metric") between states s_1 and s_2, that expresses the "distance" $d(s_1, s_2)$ between these states as a number. This expresses our intuition that it will take more time or effort to bridge a larger distance than it takes to bridge a smaller distance.

The simplest distance metric is equal to *the number of distinctions or differences between states*. This metric corresponds to the minimal number of elementary changes that one has to apply to transform s_1 into s_2.

Example: the state left–above–in front, is located at a distance 1 of the state left–above–behind, but at a distance 3 of the state right–below–behind.

The inverse of distance is *similarity*: the closer two systems are together in the state space, the more they resemble each other, that is, the less differences they display.

Application: if you compare the genetic codes (DNA strings) of two individuals, you can count the number of differences. This corresponds to the minimal number of mutations (elementary modifications in the DNA) that are necessary to change the one DNA string into the other. This gives us an indication of the time required to allow the one form to evolve out of the other, or—more precisely—the time needed to allow both to evolve from a common ancestor. In this way you can for example determine when humans and chimpanzees last had a common ancestor (about 7 million years ago).

In the case of properties with more than two values, you need to define more complicated distance metrics, such as for example the traditional Euclidean or vector distance. For three dimensions, it looks as follows:

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2}$$

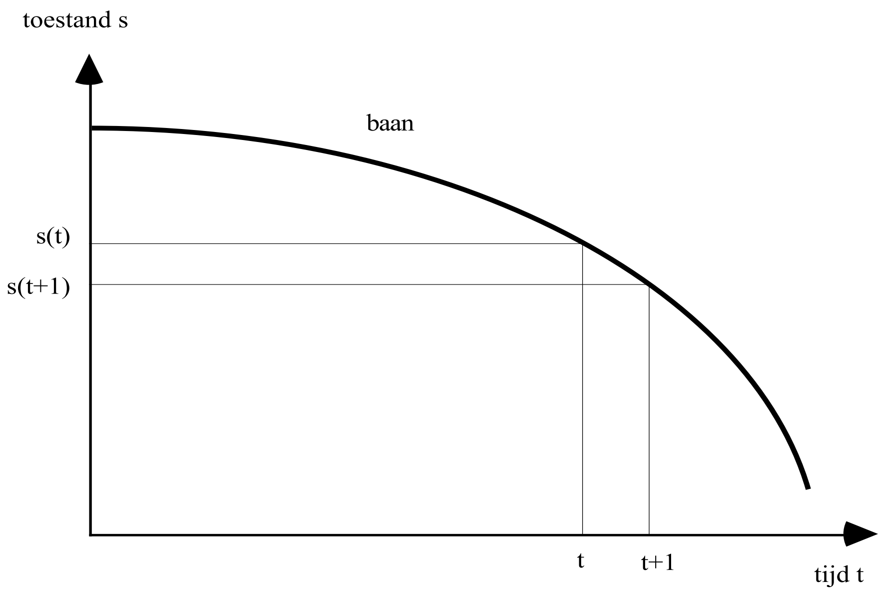
Application: if you search for information on the Internet, it is useful to know to which extent two texts are similar. For example, when you have found an interesting document via a search engine (e.g. Google), this allows you to find similar documents. The search itself is also based on similarity, in this case between the text that you enter into the search box (consisting of the word(s) that best represent your interest, e.g. "complexity evolution") and the text of all available documents. The best results for the search are those documents that best resemble the query, i.e. that contain the greatest frequency of your search terms. One way to address this problem is by defining a state space of all possible texts, where a state corresponds to a list of the frequencies of all possible words in the text. Every word determines a property or dimension, and the relative number of times that that word appears in the text determines the value of that property for this text. By calculating the distance (or sometimes the angle) between the vectors that represent two texts, you can determine how similar or different these texts are. In this way, vector distances make it possible to quickly find information.

8.7 Dynamic systems

The primary application of state spaces is the prediction of change, i.e. the determination of how the state changes through time. At every point of time t, the system is located in a specific state $s(t)$. From this state, it can make a *transition* to one or more other states $s(t + 1)$ (or possibly remain in place). Which new state is chosen depends on the *dynamics*: the whole of the "forces" that influence the system and cause it to evolve. Such dynamics is generally represented using a so-called differential equation, which expresses the speed of the variation ds/dt of the state over time as a function of the present state. Such a differential equation assumes that the state space is continuously infinite, which makes things more complicated. A simpler form, applicable to discrete or finite state spaces, is the difference equation. This indicates the difference between the state $s(t)$ and the subsequent state $s(t + 1)$, and typically has the following form:

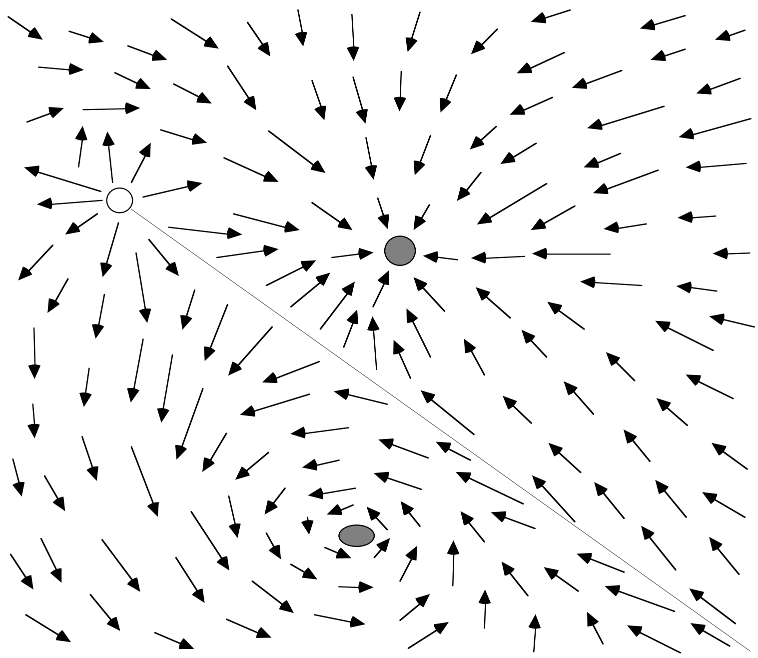
$$s(t + 1) = s(t) + f(s(t))$$

Here, f is a certain function of the current state $s(t)$ that represents the different "forces" that cause this state to move to a new state. Since this book emphasises concepts, rather than mathematical techniques to make calculations, I will not further examine these functions and equations. Suffice it to say that there are very sophisticated methods that can thus determine the evolution of a state for systems that are not too complex, and for which we exactly know all objects, properties and forces. Such a system, of which we can perfectly predict the further movement through the state space starting from the current state, is called a **dynamic system**.



We will here restrict ourselves to an overview of the qualitative properties of a dynamic system and their movement through the state space. From the newly-reached state ($s(t + 1)$, it will again move to the next state $s(t + 2)$ and so on. Such a sequence of transitions can be represented as a path or **trajectory** in the state space. This goes from earlier to later states via the present state. A trajectory is formed by the sequence $s(t)$ of all subsequent states in time t of the system. For example, a canon ball that has been shot follows a specific curve trajectory the air, depending on the force of the explosion, the orientation of the canon and the influence of gravity. A trajectory can be represented in a state space through a more or less curved or sinuous line that moves from past to future states via the present.

The basic assumption behind a dynamic system is that the next state is completely determined by the present or initial state. This means that through a certain state, there can only be one (future) trajectory: there is after all no doubt or uncertainty possible about the next state. This allows us to represent the dynamics or potential evolution of the system in a convenient manner. If we represent the state space on a piece of paper, with every point representing a state, we can in this state note in which direction the trajectory will move through that state by means of an arrow (see illustration). All arrows together then show visually how evolution occurs in different regions of the state space. We can for example immediately see whether the trajectories through two adjacent states are moving towards each other or away from each other. Sometimes we see more complicated patterns such as spirals or loops, where the trajectories circle around a central region. Such a visual representation of a dynamic system is called a **phase portrait**.



Of special interest are the regions where the arrows point inwards, but not outwards. This means that a trajectory can end up in this region from another part of the state space, but once inside, there is no trajectory leading out again. The core of such a region is called an **attractor**, since it appears to pull the trajectories towards itself (the grey areas in the illustration above). The existence of attractors makes it much easier to predict the further evolution: after all, it suffices to know that an initial state is located in the **basin** of an attractor (the region around the attractor from where all arrows and trajectories lead towards the attractor), to know that the system will necessarily also end up in that attractor. We will discuss these concepts in more depth later, when we will introduce the related visual representation of a fitness landscape. A repulsor, on the other hand, is a region from where all arrows lead away (the white circle in the illustration).

Chapter 9. Information and Entropy

9.1 Introduction

Entropy and information are no doubt the most important quantitative concepts in complexity science. As we have shown before, complexity itself is a qualitative concept that cannot be measured in an unambiguous way, because it combines the opposite poles of order and disorder, and of distinction and connection. However, if we only look at distinctions, we can determine an unambiguous metric, called **variety**, which measures the total number of distinctions in a model. Although variety is used infrequently in science (the concept was introduced by cyberneticist Ashby), it is a very simple and intuitive concept that offers a good basis from which we will deduce the better-known, but also more complicated, concepts of constraint, entropy and information, through step-by-step generalisation.

In the new, systemic worldview, information and entropy play the central part that matter and energy played in the old, mechanical worldview. Many scientists therefore tend to postulate **information** as the third fundamental substance—after matter and energy—from which the universe is constructed. However, this representation is misleading and causes a lot of confusion, because information is not a "substance". To begin with, information, contrary to matter and energy, is not conserved: it can disappear (for example if we forget something), or be created (for example if we observe something). Moreover, information is always to a certain extent subjective: it depends on the observer, for whom a certain signal can be informative or not, depending on the observer's purpose and prior knowledge. Information, like the distinctions and connections that it measures, is fundamentally relational and thus to a certain extent relative.

Entropy and information are subtle concepts with different facets. They are therefore often misinterpreted, even by scientific experts. **Entropy** is usually seen as a measure for disorder in a system, but as we will see, it is a bit more subtle than that. Entropy owes its reputation to the notorious second law of thermodynamics, which states that in a closed system entropy can only increase over time. Thus this law was the beginning of the end for the Newtonian worldview, which states that after all every change is reversible, and therefore that entropy would need to remain constant. We will discuss this law in more detail later, but to do so properly, we first need to clarify the meaning of entropy.

At first sight, **information** seems a more self-evident concept, since it is the basis of the information technology that we are by now all familiar with. Most computer scientists have however forgotten the origin of the concept, as well as the fact that its definition is directly based on that of entropy. All subtleties and misunderstandings connected to the concept of entropy are thus also found in the concept of information. The confusion is so great that it sometimes happens that one scientist proposes a definition that is the opposite of another scientist's definition. We will try to avoid this problem of the "sign" (positive or negative) by defining information as a difference that can pan be either positive or negative, depending on its starting point. More generally, we will try to create clarity in this crucial but confusing matter by continuing to build on our ontology of distinctions. This brings us directly to the concept of **variety**.

9.2 Variety

The **state space** is, as we have seen, the space of all possible states that the system can reach. The larger the state space, the more possibilities, “space” or “freedom” the system has at its disposal, but also the larger the variety of states or manifestations that the system can adopt.

Variety is a measure for the *size of the state space*, or the number of distinct possibilities. Variety V is defined as the logarithm to the base 2 of the number of elements $| S |$ of the state space S:

$$V = \log_2 | S |$$

Reminder: the logarithm of x to the base 2 is the inverse of the function 2 to the power of x, that is, $\log_2(2^x) = x$.

It follows that if $| S | = 2^N$, then $V = N$.

V is therefore equal to the *number of binary distinctions* (see 8.5).

V is however also useful for properties that are not binary. For example, for one property with five possible values, $V = \log_2 5 = \pm 2.322$.

Note: V will have to be defined differently if the number of values is infinite, because the logarithm of infinite is still infinite. Simplest case: V = number of dimensions (“degrees of freedom”).

The unit of variety is the **bit**. If a system has a variety of one bit, this means that the system has exactly *two* states or possibilities (“1” or “0”, “yes” or “no”) to choose from. A variety of N bit means 2^N possibilities.

- Why do we use logarithms?

The essential property of the logarithm is that it reduces a multiplication to an addition:

$$\log (a \times b) = \log a + \log b$$

As we have seen in 8.5, the number of states in a state space is equal to the product of the number of values for each property, or to the number of states for each object. The variety of such a “product” state space is therefore simply the sum of the varieties for each of the state spaces. This means that if you add objects or properties to the model, you only have to add the variety of those new elements to the rest. The fact that you can work with sums makes the calculations a lot easier.

Example: if one billiard ball has a variety V, then a billiard game with two billiard balls has a variety of $V + V = 2V$.

- if a billiard ball has variety V_b and a cue has variety V_c, then a game consisting of one ball and one cue has variety $V_b + V_c$
- with 3 balls and 2 cues, the variety is: $3V_b + 2V_c$

9.3 Constraint

Constraint is the opposite of freedom: it is that which reduces the number of options or possibilities. This restriction of freedom is in general not enforced from the outside, but intrinsic to the system.

Definition: there is constraint on a system if all conceivable combinations of properties are not actually possible in practice. This means that the factual set of possible states is only a part of the total state space.

- Example: billiard with one ball

The state space for the ball consists of all possible positions on the billiard table. Add constraint: place a beam in the middle of the table, so that the ball can only move in the left part. The ball is "constrained" to remain on the left side; it is "restricted" or "confined" to the left part. The number of possible states has been halved. The variety has been reduced with 1 bit (logarithm of the factor 2).

- Example: state space of a berry

Let us assume that the model of a berry distinguishes two binary properties: *colour* with values "red" or "green" and *size* with values "large" or "small". Two properties with two values each means $2 \times 2 = 4$ states and therefore $V = 2$ bit. In practice, however, all green berries turn out to be small (because they are unripe) and all red berries large. This means that only two states are possible in practice: (green, small) and (red, large). This means that $V = 1$ bit. There is thus constraint or restriction that excludes or forbids the combinations (green, large) and (red, small).

- Example: billiard with two balls

The state space for two balls is the product of the two individual state spaces, i.e. the set of all pairs of the form (position 1, position 2), where position 1 denotes an arbitrary position on the table for ball 1. Now add constraint: stick the balls together. The first ball can still take an arbitrary position on the table. The second ball on the other hand becomes extremely restricted in the choice of remaining positions, since it necessarily needs to stay at the same distance to the first ball. The state space for two balls stuck together is much smaller than that of two independent balls.

Constraint C is defined as the maximal variety minus the actual variety:

$$C = V_{\max} - V$$

This means: C represents the loss in variety or freedom relative to the best conceivable situation.

Example: berries: $C = 2 - 1 = 1$ bit; billiard table divided in two: $C = 1$ bit

Constraints may seem negative for the system, since it restricts its freedom, but it is in general positive for the observer. The observer will after all know better *where* the system is located, and will therefore have more control over it. With the example of the berries, it suffices to know that a berry is small (for example by grasping it), to be able to deduce that it is also green, even if it is dark and you cannot distinguish the colours.

Constraint on a system consisting of several objects will generally lead to a relationship or dependence between the objects. For example, with balls that are stuck together, the position of the one ball is dependent on the position of the other. In the example of magnetisation, the direction of each magnet is determined by the direction of the others. Constraint can thus be seen as a measure of order: mutual dependency or connection.

9.4 Entropy

Imagine that we do not know the precise state s of a system, but only the probability $P(s)$ that a system would be in state s. This is normally the case for systems with a large number of components, where we cannot determine every property for all components. In that case, it is useless to determine the variety by adding up the different possible states, since different states will generally have different probabilities. However, we can easily generalise the concept of variety for this more complex situation, leading to what is called the **entropy** H. For the probability distribution $P(s)$, this is defined as follows:

$$H(P) = -\sum_{s \in S} P(s).\log P(s)$$

Thus, entropy is calculated as the sum over all states of the logarithms of the probabilities of these states, multiplied with the same probabilities. The logarithm is still defined to base 2. A minus sign is added to the sum, because the logarithm of a number smaller than 1 (which is by definition the case for a probability $P(s)$) is negative. Otherwise, the sum would be negative. Although this formula looks complicated and not very intuitive, it is so important that we have to investigate it.

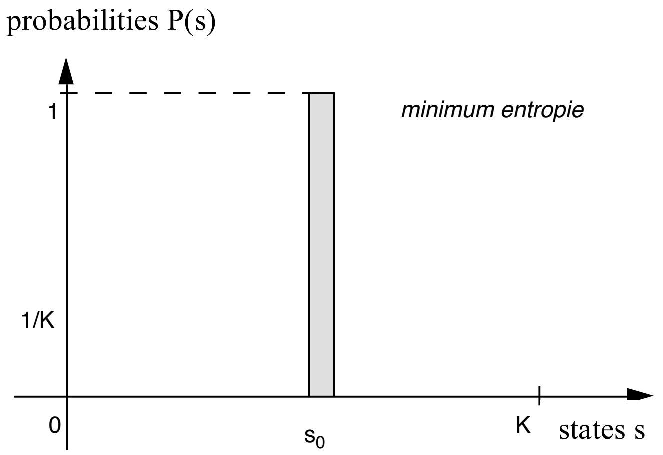
Entropy is a measure for **uncertainty**, or our lack of knowledge regarding the state of the system: the less we know, the greater the entropy. This is best clarified by means of some special cases.

- Minimal entropy:

If we are certain that the system is located in state s_0, i.e. $P(s_0) = 1$, $P(s \neq s_0) = 0$, then:

$$H(P) = 1 \cdot \log 1 + 0 \cdot \log 0 + 0 \cdot \log 0 + ... = \log 1 = 0.$$

This is minimal entropy or no uncertainty. In other words, we have full knowledge or information about the state. If we draw the corresponding probability distribution (below), we see one peak for s_0, where the maximum value 1 is reached, while all other states remain zero.



- Maximal entropy:

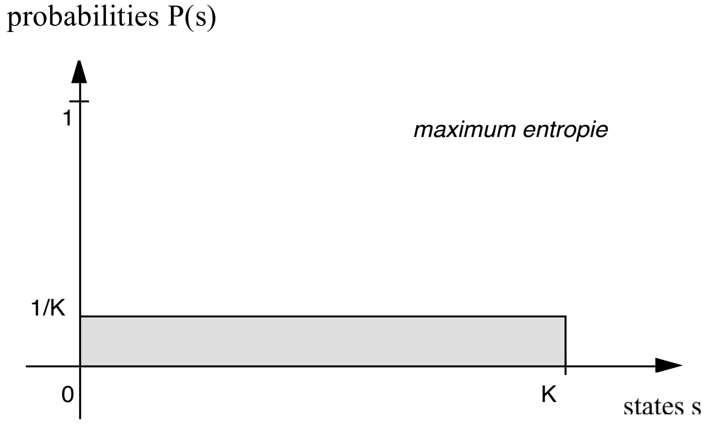
Now imagine that we have no indication at all that a certain state s_i would in any way be more or less probable than any other state. This means that all states have the same probability:

$$P(s) = 1/K, \text{ with } K = |\,S\,| = \text{number of states in state space } S.$$

Then:

$$H(P) = -\,K \cdot (1/K) \cdot \log\,(1/K) = \log K = V.$$

That means that the entropy is reduced to the logarithm of the number of states, or the (maximal) variety V. Here, entropy reaches its maximum. This is the situation where we have absolutely no knowledge or information about the state of the system, and where our uncertainty is therefore the largest. When drawn this gives a completely flat, homogenous probability distribution, where all states have the same value $1/K$.

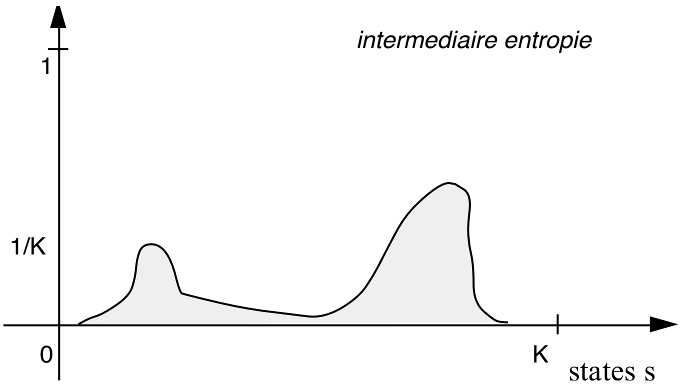


- Intermediate entropy:

In the intermediate case, i.e. probabilities that are different from each other, but that do not reach the maximal value 1, entropy H has an value intermediate between the two previous values:

$$V > H(P) > 0.$$

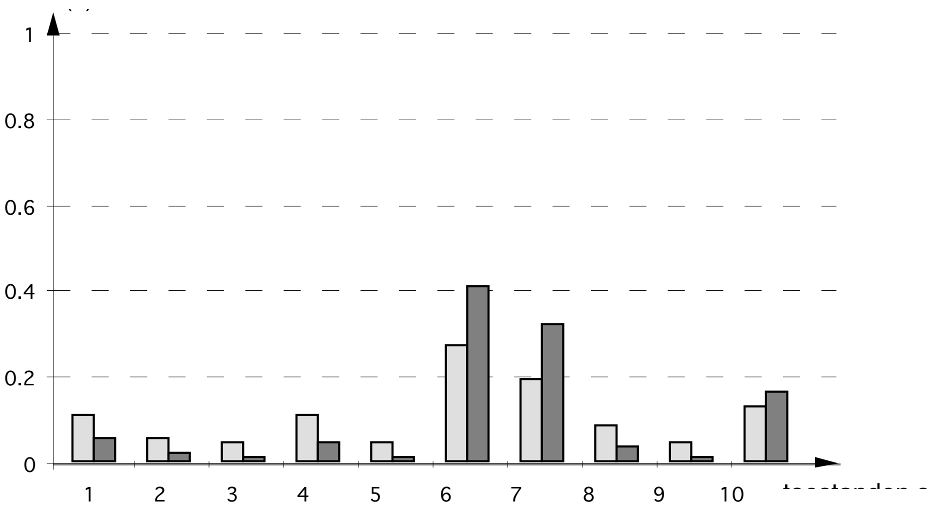
probabilities P(s)



states s

The more even or *homogenous* the probability distribution (i.e. the less differences there are between the probabilities for the different states), the larger the entropy H. The more *heterogeneous* the distribution (i.e. the higher and narrower the “peaks” of states with a high probability relative to other states), the lower the entropy. In the figure below, two probability distributions in a state space with 10 states are placed side by side. For the more “even” distribution, shown with light-grey columns, $H = 3.04$; for the more “peaked” distribution, shown with dark-grey columns, we find the smaller value $H = 2.33$. The maximal entropy is: $H = \log_2 10 = 3.32$)

probabilities P(s)



states s

- Interpretation

Entropy is a direct generalisation of variety. This means that it can be understood as a measure for the freedom of the systems to adopt different states. It can also be seen as a measure for our lack of knowledge (uncertainty) about the system. Thus we can say very little about a system with high entropy.

Entropy can also be seen as a measure of disorder. Freedom for the system means independence of its components, and therefore distinction, but no connection. Even if we would know the state of a part of a system with high entropy, we would in general still know little about the other parts. Conversely, this means that a system with low entropy could be seen as ordered: since we already know almost everything, it is perhaps enough to know the state of a small part to also be able to determine the state of the rest.

Note that this interpretation in terms of order assumes that there are connections between the parts of the system, which are responsible for decreasing our uncertainty. Until now, however, we have only paid attention to the distinctions in our formalism of the state space, and the connections therefore do not play an explicit part in the definition of variety and entropy. This is a shortcoming of the representation of a system in terms of state spaces, which therefore makes the (currently used) interpretation of entropy as disorder precarious.

- Important note: thermodynamical entropy

The above is a definition of so-called "statistical" entropy (based on probabilities). It was introduced by the physicist Boltzmann, and generalised by the cyberneticist Shannon. There is however also "thermodynamic entropy" (cf. 3.1), which is a measure of the spread or **dissipation** of energy in the form of heat. In many cases (as originally foreseen by Boltzmann), statistical and thermodynamical entropy coincide, because dissipation means that the distribution of energy becomes more homogenous. Statistical entropy is however a more general concept: it can also be used in models in which heat or energy have not been defined. Statistical entropy depends on our knowledge, and can therefore change when our knowledge increases or decreases, even while the thermodynamical entropy does not necessarily change.

9.5 Information

Information is that which cancels our lack of knowledge or our uncertainty. Information can therefore be defined as *decrease in entropy*. Suppose that a system initially has entropy H(before), and that we get information I about the system (for example by observing it, or through a tip that someone gives us). This leads to the new, lower entropy H(after). The received information is:

$$I = H(\text{before}) - H(\text{after})$$

<u>Special cases:</u>

- Suppose that H(after) = 0 (we now have certainty about the state of the system). Then: $I = H$ (before).
- Now suppose that H(before) = 0 (we *were* certain about the state of the system). Then $I = -H$(after), in other words the given information is negative; we have lost information.

These special cases explain why some authors define information as equal to entropy, and others as the opposite of entropy. We can however not define information in se, like we defined entropy and variety. Information indicates the result of a process (typically an observation or communication process) that changes our knowledge. This implies the comparison between two situations, before and after, or with and without, the provided tip.

Just as for variety, constraint and entropy, the unit for information is a **bit**. This is the same bit with which the memory capacity of a computer is measured, or the transmission speed of a communication network. This formula for information and the bit measure for information transmission were developed by Shannon to enable the measurement of the capacity of communication channels (e.g. telephone lines). Below are the definitions of some common units of information storage or transmission that are derived from the bit:

1 Byte = 8 bit (typical length of one character sign in the ASCII alphabet)

1 Kilobyte (KB) = 1000 (or more precisely $1024 = 2^{10}$) Byte

1 Megabyte (MB) = 1 million Byte

1 Gigabyte (GB) = 1 billion Byte

10 Mbps = 10 million bit per second

- Examples:

The answer to a binary "yes-no" question (for example, "Is it raining?") provides 1 bit of information—assuming that the two answers have the same probability.

If the probabilities are different, the obtained information is however smaller than 1 bit. For example, imagine that you ask the question "Is it raining?" in the Sahara. The probability that the answer to that question will be "yes" is very small. In most cases, the answer "no" will only confirm your expectations and hardly give any information at all. Only in the exceptional case that the answer is "yes", did you really get significant information. Suppose $P(\text{no}) = 0.99$ (i.e. there is a 99% chance that it is not raining), and therefore $P(\text{yes}) = 0.01$, then:

$$I = -\,0.99 \cdot (\log_2 0.99) - 0.01 \cdot (\log_2 0.01) = 0.08 \text{ bit}$$

This can also be seen as the average (weighted according to the probability) of $I(\text{no}) = -\log_2 0.99 = 0.014$ and $I(\text{yes}) = -\log_2 0.01 = 6.64$. $I(\text{no})$ is therefore much smaller than $I(\text{yes})$, but because "no" occurs much more frequently than "yes", it weighs much heavier in the information you obtain.

The answer to a question with more than two possible answers generally gives you more than 1 bit of information. For example, if the possibilities are "sunny", "cloudy", "overcast", "showers" and "thunderstorm", each having the same probability, then the weather forecast gives you $\log_2 5 = 2.322$ bit of information.

- Application: compression of data

Most data (for example in natural language) contain far less factual information than you would expect if you would count the number of letters or characters used to express these data (on a computer, 1 character typically corresponds to 1 byte). The reason is that the probability of the occurrence of letters in a language is not at all homogenous. For example, in English, "e" occurs more frequently than "x", "er" appears more frequently than "yr" and "the" appears more frequently than "hte". In principle, a text in such a language can be recoded so that the same information is presented with fewer letters. The most efficient code is the one where every character would have the same probability of occurrence, because then the original entropy $H(\text{before})$ is maximal (homogenous probability distribution), and the same will then apply to the obtained information I for a message of a certain length.

Concretely, we could achieve this by replacing frequently occurring combinations of letters by one single, new character. This would greatly reduce the number of required characters. To prevent having to create too many new characters (which would increase the number of distinctions and therefore the entropy), we can replace rarely occurring letters or combinations by a combination of other, existing signs. This will make the text a bit longer again, but since this by definition happens rarely, it will not matter a lot.

This kind of recoding is called *compression*, because it reduces the "space", measured in bits, that data take up in the memory of the computer or during transmission, so that more information can be stored or transmitted. It is because of compression that we can download complex files such as pieces of music, photos and movies from the Internet without having to wait too long and that we can store them on disks with limited capacity. The theoretical problem with compression is that we do not actually know the most efficient way to compress: if we search a bit further, we can always find a shorter, even more compressed coding. This is the same problem that I have already mentioned with the attempts to define complexity as the length of the shortest possible description.

On the other hand, it is not really recommended to find the most efficient coding. A non-efficient coding has what is called *redundancy*: it contains more characters than necessary. But that also means that it does not really matter if a sign is accidentally left out or misread during transmission. Redundant codes (such as natural language) have the advantage of robustness: mistakes can in general easily be corrected, because the remaining, correct characters still provide us with enough information to infer the missing or incorrect characters correctly. Even a txt full of speling missteaks can be nderrstod acccrastely.

Interpretation: Information can again be interpreted as a measure of constraint or order: the more information we have about the state of a system, the fewer possibilities remain for the system.

Note: Information or entropy can also be used to measure the *mutual dependence* of components. For this, one does not start with $P(s)$, but with $P(s_a|s_b)$. This is the conditional probability that component a is in a state s_a, given that b is in state s_b.

This defines $H(a|b)$, i.e. the uncertainty about the state of a, given the state of b. The smaller $H(a|b)$, the more certain we are about the state of a if we already know the state of b, that is to say, the more information b gives us about a. $H(a|b)$ can thus be seen as a measure of the (absence of) connection between b and a.

9.6 Limitations

To what extent can variety, constraint, entropy or information be used to measure complexity? Variety and entropy measure the degree of disorder or indeterminacy. This is sometimes called "disordered complexity", but is not really complex. Constraint and information measure the degree of order or definiteness. This can perhaps be seen as a primitive measure of dependence, but not of complexity.

As a matter of fact, complexity rises only if both entropy and information rise. This seems paradoxical, since increase of the one has been defined as decrease of the other. Still, we can imagine a situation in which both in a sense increase. This does however require the expansion of the state space with additional distinctions (objects and/or properties). A larger state space implies greater variety / entropy. At the same time, the connections (relational constraints) between the components (objects or properties) have to increase. However, to describe these kinds of processes, we also need a distinction dynamics, that is, a theory that explains how distinctions can increase or decrease over time.

Chapter 10. Variation and selection principles

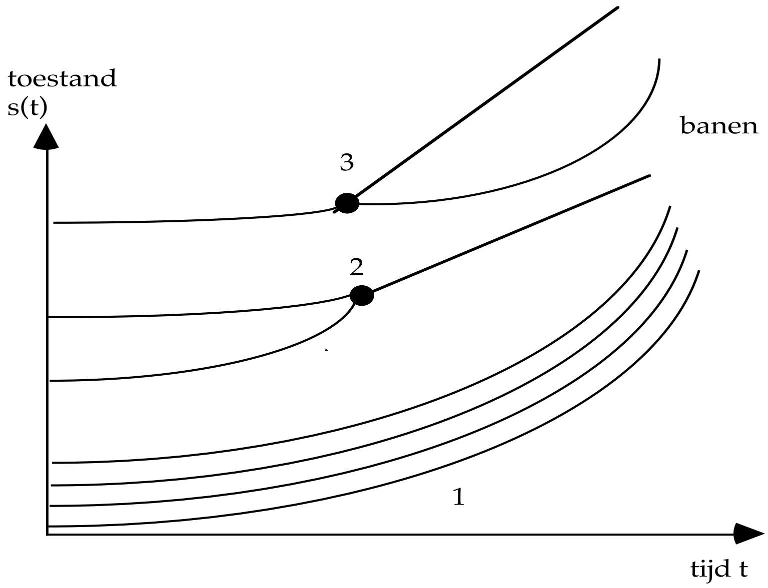
10.1 The dynamics of distinctions

We are now interested in the evolution of a system over time. We have seen in the section about dynamic systems that this can be predicted using a **trajectory** through the state space. This is a classical way to describe change, which can be traced back to Newtonian mechanics. The question that we now ask however is: how do complexity and the related phenomena of variety, entropy and information change? To answer this question, we have to extend the Newtonian view of trajectories, and allow for trajectories that are not uniquely determined.

- Newtonian dynamics

In Newton's mechanics, every state is unambiguously followed by a unique other state. There is no uncertainty or unpredictability, since the initial state (cause) completely determines the following state (effect), and vice versa. This implies that the trajectories through the different initial states will always remain parallel: they do not intersect. Otherwise, when you hit an intersection (**bifurcation**, see 3.7), further movement would not be predetermined, since the state at the intersection point would have the choice between two different trajectories.

Therefore, all distinctions are conserved: two different initial states correspond to two different final states, and vice versa. Variety, entropy and information are therefore constant too. Our knowledge about the initial state is no larger or smaller than our knowledge about the final state. *Information is retained*: it neither increases nor decreases. We can summarise this form of dynamics or evolution as **distinction conservation**.

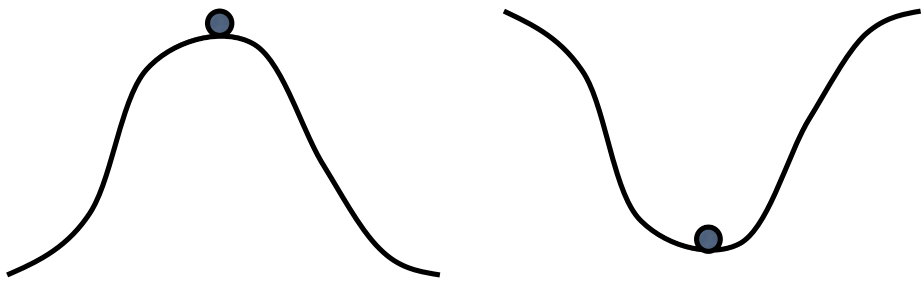


1 distinctieconservatie
2 " destructie
3 " creatie

- Attractor dynamics

<u>Example:</u> a ball rolls into a hole. Whatever its origin, it will always roll towards the deepest point and come to a stop there. The state at the deepest point thus plays the role of an **attractor**. The ball cannot get out of the hole on its own: the movement is irreversible.

Here, distinctions are erased. Indeed, different initial states or trajectories all come together in the same attractor or final state. Variety decreases, information increases. We did not know where the ball was initially, but we are now certain that it is in the deepest point. This is a form of **selection**: of the different possible states, most are eliminated and finally one remains. Selection can therefore be seen as the dynamic equivalent of constraint: through this process, the state space is reduced to a subspace (in this case with only one state). We can summarise this kind of evolution or dynamics as **distinction destruction**.



- Stochastic dynamics

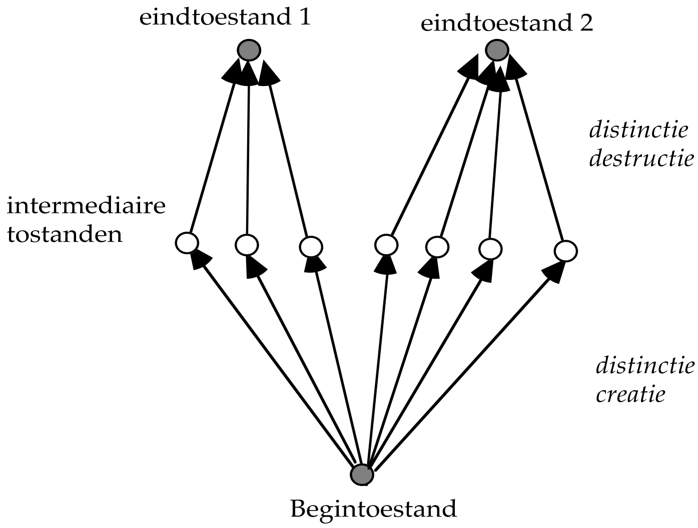
<u>Example:</u> a ball rolls down from an unstable position at the top of a hill. We do not know in which direction it will roll down, and cannot predict where it will end up. The process is stochastic, which means non-deterministic.

Here, distinctions are created. The same initial state will indeed give rise to a variety of possible final states. If the experiment is repeated, the outcome will always be different. Variety or entropy increases, and information decreases. Initially, we knew exactly where the ball was, afterwards we no longer know. This is the equivalent of **variation**: the creation of additional possibilities. We can summarise this type of evolution / dynamics as **distinction creation**.

- General evolution

A general process will be a combination of subprocesses characterised by distinction creation (variation), distinction destruction (selection) and distinction conservation (causal evolution). Certain distinctions will be conserved / created / destroyed at certain moments, depending on the forces that act upon the system.

If we start from a single state (one distinction), this single state can only be retained or split into different possible states (distinction creation). As soon as there are more distinctions (states), these can again merge, as a wholly or in parts (distinction destruction). This is a general scheme for an arbitrary process that starts from a certain initial state, in which we can recognise the consecutive phases of distinction creation and distinction destruction (see illustration).



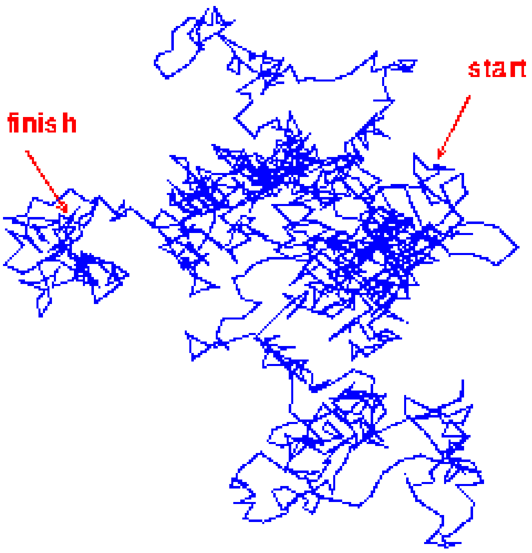
In practice, both phases will run together constantly, but by separating them conceptually, we may hope to get a better insight into the components of the process. A theory ("dynamics") of such processes should give us an idea of the "forces" or "mechanisms" that govern these phases. We will now discuss the most fundamental of these mechanisms, with the help of examples and basic principles.

10.2 Variation without selection: drift

A system that is left to its own devices will in general vary randomly, under the influence of all kinds of unknown disturbances. We assume no systematic force or preference that pushes it in a specific direction. Therefore, we cannot determine the trajectory in the state space in advance. This means that we *lose information* about the system: if we initially knew the state of the system, then we no longer know it after unknown influences have changed that state.

- Example: a bottle thrown into the sea

The bottle is subjected to the waves, currents, wind, etc. The longer the bottle is adrift, the further it will in general be removed from the place where we threw it into the sea. We however do not know in which direction the bottle will move, or where it will end up. The longer we wait, the larger our uncertainty about the location of the bottle becomes.



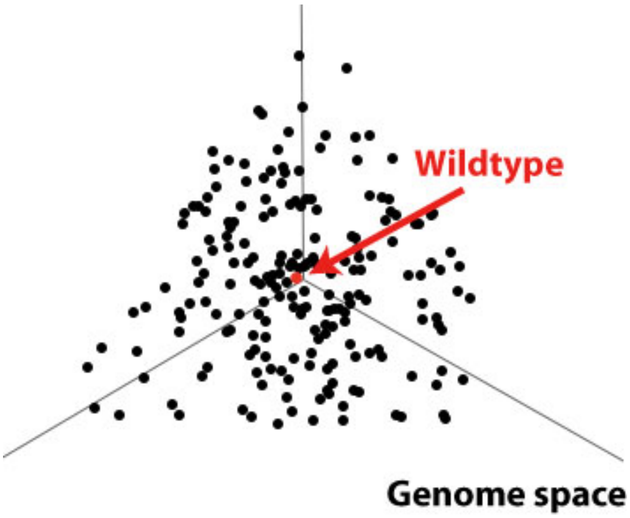
- Example: Brownian movement

A miniscule dust particle in a stationary fluid will still move in an unpredictable manner. This can be seen under a microscope. The explanation is that fluid molecules collide with the particle from random directions. Every collision causes the particle to move a bit.

This type of movement is also called a *random walk*. It can be compared with the trajectory of a drunk who staggers around without goal or direction. In the illustration to the left, you see a simulation of a random walk, from the starting point to the point where the simulation was finished.

- Example: neutral evolution

A population of animals and plants can evolve even without natural selection. By accident, there will sometimes be more animals born with gene A than with gene B. If it happens that no animals with gene B are born, this gene will permanently disappear from the population. This is called **genetic drift**. In this way, a species can evolve independently of its environment.



If the genes also undergo accidental mutations that do not confer an advantage or disadvantage for survival, the process is called **neutral evolution**: the genetic makeup, and therefore the organisms themselves, change over time, but this evolution is not controlled by natural selection. This happens often with bacteria and viruses, which after all mutate easily.

The effect can be visualised in a state space: a "wildtype" (initial state) that multiplies will produce descendants that are all slightly different from their ancestors, albeit in a random way. The more generations, the more mutations, and therefore in general the more differences. If we indicate these descendants with points in a state space, you will first see a concentrated "cluster" around the wildtype, which gradually fans out to a fuzzy "cloud" which covers an ever-larger region. Because it is difficult to draw dividing lines or boundaries that distinguish the one kind or variant from the other in such a cloud, such a distribution of related, but different types is called a **quasi-species**.

10.3 The second law of thermodynamics

The previous observation about the spontaneous decrease of information under the influence of random variations can be explicitly formulated as a law:

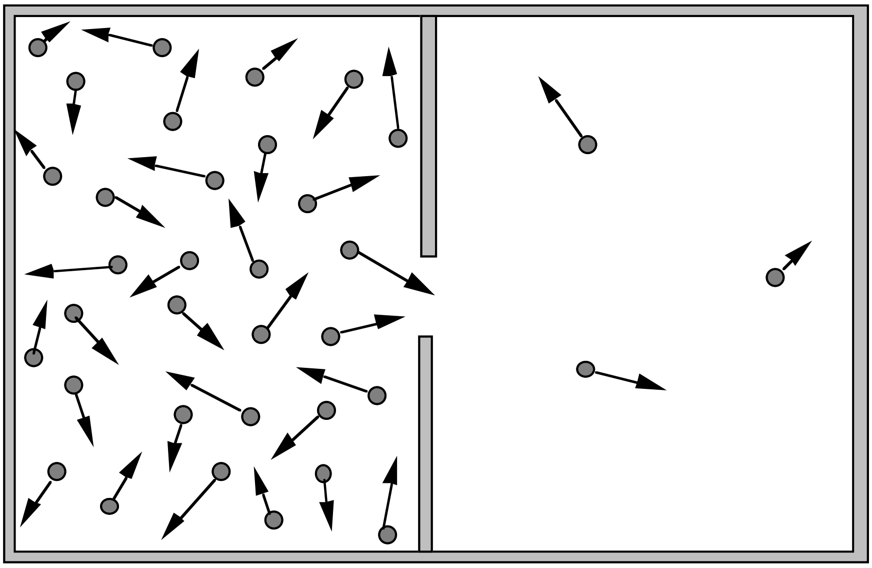
> *In a system without selection, entropy can only increase or remain the same, but never decrease.*

Entropy or uncertainty aims in a certain sense to reach a maximal value. Only when it has attained this maximum will evolution stop and will the system reach equilibrium. This means that such evolution is **irreversible**: entropy can never decrease again to its original value. This is an alternative formulation of the famous second law of thermodynamics. Let me illustrate this law with a classical example.

- Example: the box with two compartments

Consider an airtight box with two compartments that are separated by a wall. In the left compartment there is a gas—i.e. a collection of molecules that move in different directions. The compartment on the right is initially empty. Now imagine that an opening is made in the dividing wall. What happens? The gas spontaneously flows into the empty compartment, until both compartments are homogenously filled. After that, nothing changes anymore: the system has reached equilibrium.

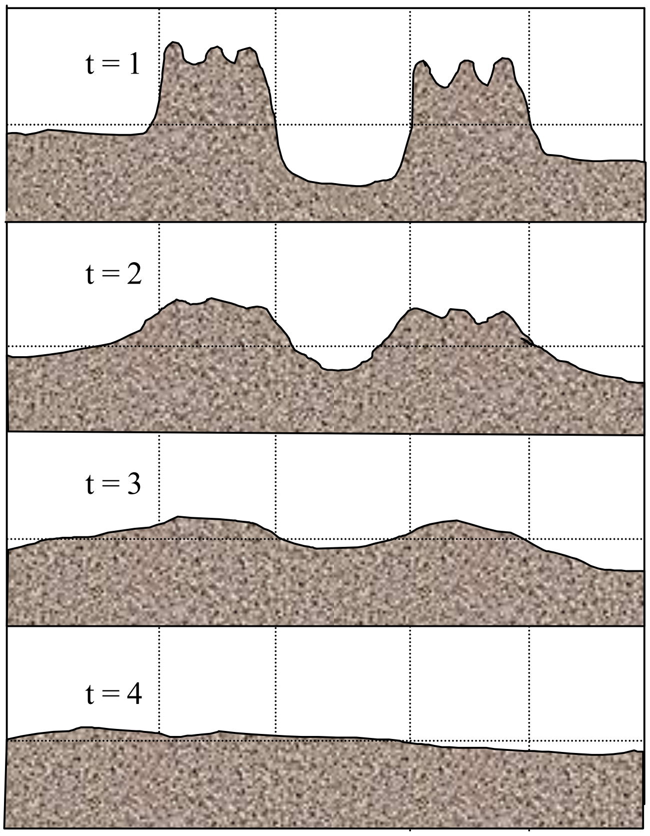
Entropy has clearly increased. First, we were certain that an arbitrary gas molecule would be found in the left compartment: $P(\text{left}) = 1$, $P(\text{right}) = 0$. Afterwards, we have a fifty-fifty chance to find the gas molecule either left or right: $P(\text{left}) = P(\text{right}) = 0.50$. Our uncertainty about the location has therefore increased, while the constraint on the molecules has decreased: they are no longer restricted to the left compartment.



<u>Explanation:</u> A molecule in the left compartment that moves to the right, at the level of the hole, will end up in the right compartment. A molecule on the right compartment that moves to the left, at the level of the hole, will end up in the left compartment. The probability of this happening is the same for the left and for the right. Initially there are however many more molecules on the left side, and therefore many more molecules that will fly through the hole to the right. Thus, on average more molecules will move from left to right than vice versa. *This is why the number of molecules on the right will increase*. However, when both compartments have become equally full, the average number of molecules that move left → right and right → left will be equal. That is why the number of molecules left or right now remains the same.

This same mechanism takes place in all systems whose components are not distributed homogenously and which are subject to random movements or fluctuations. A larger number of components will leave the more populated regions to end up in the less populated regions, than vice versa. The inhomogeneity, and the accompanying structure, organisation or **differentiation**, will therefore spontaneously disappear or be erased.

- Examples:
- A drop of blue ink in a glass of water spreads until the fluid gets a uniform light blue colour.
- An ice cube in a glass of water melts, and warm and cold water mix until everything has the same temperature.



A sand castle will gradually fall apart under the influence of the weather until the beach is more or less smooth again. Note that the level of the sand in the horizontal position *s* is proportional to the probability *P*(*s*) to find a grain of sand in that position: the higher the heap of sand, the more grains of sand, and therefore the larger the probability. The distribution of the grains of sand is therefore analogous to a probability distribution. The erosion and collapse of the sand castle thus corresponds to the evolution of the probability distribution from one with low entropy (i.e. peaks and valleys) to one with high entropy (i.e. flat). This becomes clear if you compare the picture on the left, which shows consecutive stages (t = 1, 2, 3, 4) of erosion of a sand castle, with the pictures of entropy distributions in section 8.4. The horizontal line here represents the equilibrium distribution, where all probabilities are the same.

- Interpretation:

In systems that are left to their own devices, disorder tends to increase. All complex systems are subject to wear and tear, that is, to the gradual loss of structure. Examples are the accumulation of dust and dirt, erosion and decay, rust, forgetting (the spontaneous fading of information in memory), deterioration and aging ending in death.

Applied to the universe as a whole, this idea appears to imply that all organisation or structure will eventually disappear. This vision is called the *heat death* of the universe, because in thermodynamics increase of entropy is accompanied by dissipation of energy in the form of heat. In the clockwork metaphor (section 1.3), the heat death is the situation in which the battery of the clockwork has run out, and there is no longer any usable energy available to allow the clockwork (the universe) to continue functioning. Everything has become uniform, grey, homogenous: there is no longer any form, change or structure; every complexity or differentiation is lost.

If you consider this interpretation as an absolute natural law, then there is no room for self-organisation, life, or progress. This produces a pessimistic and fatalistic worldview. The only remaining explanations for the life or the organisation that we see around us are then the following:

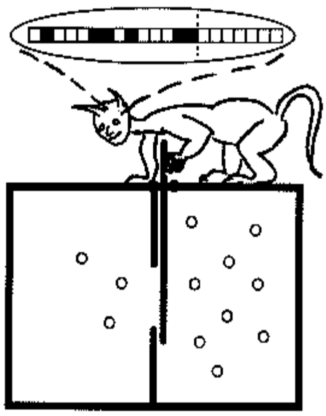
a) the result of an extremely unlikely accident (cf. the approach of the biologist J. Monod in his classical book *Le hasard et la nécessité*)
b) the creation by some supernatural force that is able to counteract the natural increase of entropy (e.g. the "*élan vital*" or life force postulated by philosopher H. Bergson)

In this book, we will defend a more realistic interpretation of the second law of thermodynamics: the increase of entropy through random variation is only one side of the coin. There is a complementary mechanism, namely selection, that has the opposite effect.

We do have to make an important observation. The traditional formulation of the second law states that entropy cannot decrease *in a closed system*. "Closed" here means that there is no exchange with the environment. This law is in general valid for thermodynamic entropy (dissipation of heat). We will now show, however, that it is not necessarily valid for statistical entropy (uncertainty). The required condition for increase of statistical entropy is not that there is no exchange with the environment, but that there is no selection.

10.4 Selection through asymmetrical preference

- Maxwell's demon

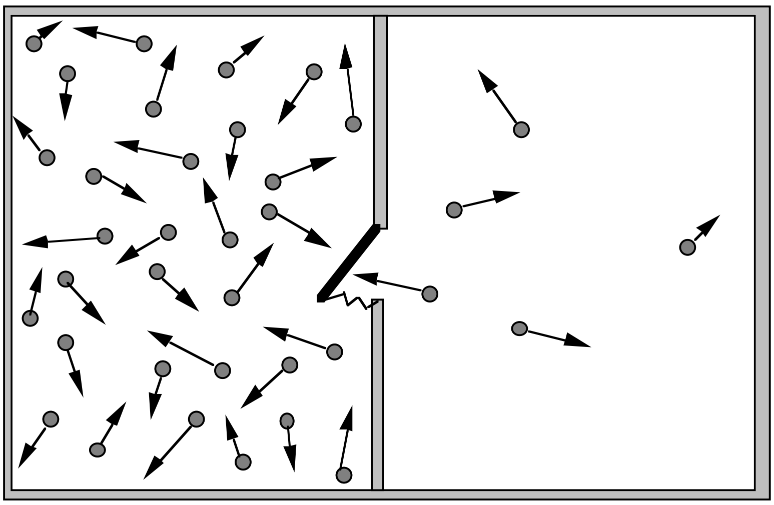


Consider again the box with two compartments, with an opening in the dividing wall. Both compartments contain the same amount of gas. Now imagine that there is a door across the opening. The 19th century physicist Maxwell wondered what would happen if there were a small demon at that door who would only open the door for molecules that were coming from the right. Thus:

- Movement left → right: door closed. The molecule remains on the left.
- Movement right → left: door open. The molecule leaves the compartment on the right and enters the one on the left.

<u>Result:</u> the right compartment empties itself into the left, and entropy decreases.

This contradicts the statistical interpretation of the second law of thermodynamics. Many scientists have therefore argued that such a demon cannot exist. We can however replace the demon with a simple mechanism: a spring that allows the door to open if the molecule collides with it from the right, but that otherwise pulls the door shut (see illustration below).



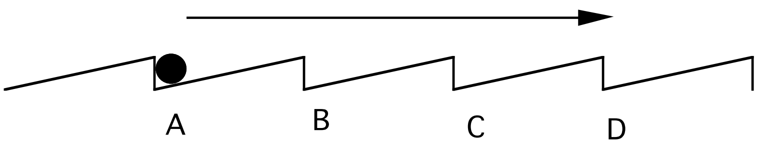
In both cases, demon or spring, the principle is the same: to decrease entropy, it suffices to apply a preference or *asymmetry* to the movements: right → left is preferred over, or is easier than, left → right.

<u>Note:</u> this thought experiment does not contradict the second law of thermodynamics in its original meaning. The *thermodynamic* entropy does increase: heat is dissipated, since the collisions of the molecules against the walls and the door release heat. However, the *statistical* entropy (uncertainty) concerning the positions of the molecules decreases.

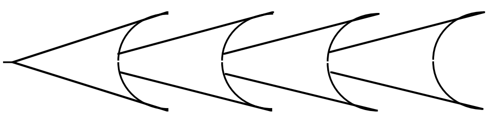
- Example: the magical carpet

This is a simple experiment that you can do yourself. Place different small, heavy objects (coins, bolts, pebbles, ...) on a loose piece of carpet. Move the carpet back and forth in both directions. Despite the random movements of the carpet, the objects appear to have a preference to move into a specific direction (for example to the right, depending on the carpet). If you keep up the back-and-forth movement long enough, they will eventually all fall off the same right end of the carpet.

The reason is that the hairs of the carpet point in a certain direction (in this case to the left), and thus impede movements in the opposite direction. On average, an arbitrary movement ("random walk", drift) will therefore be turned into a goal-directed movement. All objects are dragged by this movement to the endpoint, where they heap up. Again there is decrease of entropy: dispersed systems become concentrated in a small part of the state space.



This principle can be easily explained with the picture above. When the jagged surface is shaken randomly (movements in both directions are equally probable), the ball will still only move in the direction of the arrow: A → B → C → ... The reason is that a movement from right to left is blocked by the steep cliff, while a movement from left to right goes smoothly over the gentle slope.



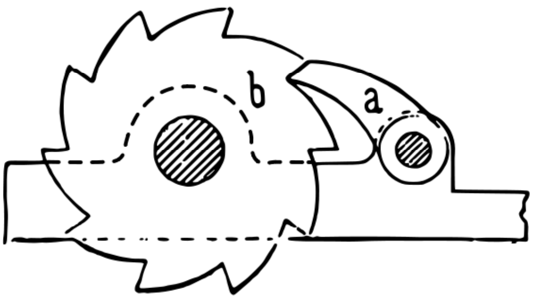
Another example is an ear of corn (depicted schematically above). This easily enters clothes (such as a woollen jumper) with its sharp point. Once it is in the clothes, however, it is very difficult to pull the ear out again, since it has barbs. The only way to get rid of this awkward thing without damaging the jumper is by pulling it further in until it has passed through the fabric. Various seeds use the same mechanism, using barbs to stick to clothes or animal fur, and thus profit from the movements of the animal to reach new grounds to germinate.

- Asymmetrical evolution

These examples illustrate a general principle of evolution: the transition from a state a to a state b is in general not as likely or easy as the transition from b to a. Transitions are usually **asymmetrical**: $a \rightarrow b$ and $b \rightarrow a$ are not equivalent. One of the two is in general preferable.

In this case, we can speak of **selection**: there is a preference for a certain direction and therefore for one of the two states. Imagine that $a \rightarrow b$ is easier than vice versa, then most systems will leave a and end up in b. If there is no other state c, for which the preference would be $b \rightarrow c$ rather than $c \rightarrow b$, then the systems will accumulate in b. State b functions as an **attractor**, a region in the state space that "attracts" the system.

- The ratchet effect



Certain movements are only possible in one direction. For example, pedalling can make a bike only go forwards, not backwards. The mechanism that enables such a unilateral movement is called a *ratchet*. A ratchet is simply a round cog (b in the illustration) of which the teeth are asymmetrical as in the previous illustration. When a catch (a) is pushed against the steep side of the tooth with a spring, this will prevent all movement of the ratchet towards the catch (clockwise in the illustration), while allowing movement away from the catch (counter clockwise).

Random variation in such a situation will be blocked in the one direction, and facilitated in the other direction. The result is that random variation is converted into goal-directed, irreversible change. This propels evolution in a specific direction: "progress", without the possibility to move backwards again. This irreversibility of evolution due to asymmetrical preferences is called the **ratchet effect**. (Whether the direction "forwards" will also be the direction of "improvement" will be studied later.)

- Conclusion

With random, unknown disturbances, all possible movements in the state space have the same probability. This leads to a diffusion from a small part of the state space to a larger part, and thus to the increase of entropy. However, in many cases, the probability of movement is not uniform. Movements in one direction are easier than in another direction: there is a selective preference. In that case, systems will accumulate in the region of the state space that is easiest to enter (and hardest to get out of). Entropy can therefore either decrease or increase, depending on the presence of selection.

10.5 Order out of chaos

Random variation produces entropy or disorder. Selection produces information or order. Together they produce *more* order. Indeed, selection blocks variations in the "wrong" direction, but allows variations in the "right" direction. Increase of random variation thus leads to:

1. more disorder when there is no selection

2. more order when there is selection

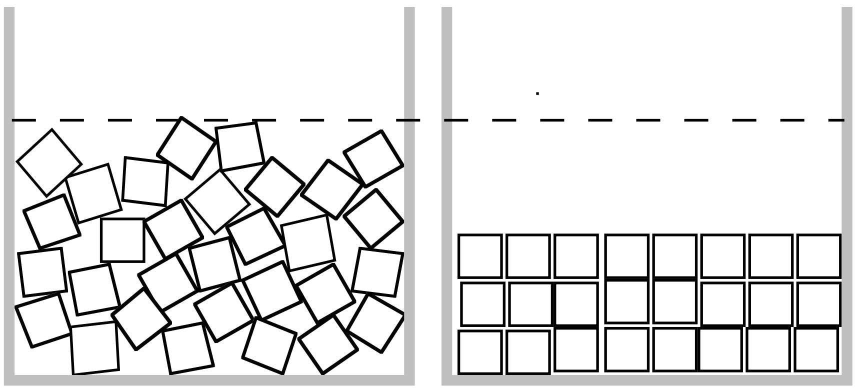
In other words, if there is a selective preference for certain configurations, *more* uncontrolled variation (chaos, noise, disturbances, fluctuations) will lead to more order. This seems paradoxical. Yet, it is a very fundamental principle with wide applicability, especially to all forms of self-organisation. This principle has been suggested by different scientists under different names, such as:

- *order from noise* (von Foerster)
- *order through fluctuations* (Prigogine)

- Example: shaking causes volume reduction

Shake a jar that you have filled with many small chunks or pieces, such as tealeaves, nails, grains of sand, or salt. The more you shake, the lower the level of the material comes to be. The jar, which initially seemed full, now appears to have quite some space left that can be filled up. The selective preference is as follows: when the jar is shaken, chunks that are closer to the bottom and closer together will have a harder time coming back up than chunks that are assembled more loosely. This is why all pieces "aim" for an as tight a concentration as possible, on the bottom. Shaking more (more variation) leads to a smaller volume (more order).

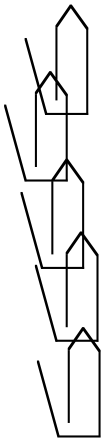
When the pieces have all exactly the same shape and size, the most compact configuration normally is the one where the pieces lie at exactly the same distance from each other, like in a crystal. Indeed, imagine that pieces *a* and *b* are at a shorter distance from each other than *c* and *d*. This implies that the distance between *c* and *d* is not minimal, because what *a* and *b* can do, *c* and *d* can in principle do as well. The optimal configuration therefore is the one in which no relationship between two pieces is different from any other relationship. This principle has been illustrated below. On the left you see an unordered heap of square blocks in a rectangular box. On the right, you see the same blocks, but now regularly aligned, with minimal spacing. It is clear that the blocks in the right box take up much less space. In the case of identical components, such as atoms and molecules, the selected configuration is thus very clearly ordered, in the sense of symmetrical or regular.



- Application: the hardening of metal

To make metal as hard as possible, one should obtain a crystal structure that is as regular as possible for the metal molecules. Indeed, a deviation from this order would imply that some molecules are farther away from each other than others. That means that they can move more easily relative to each other when the metal is put under stress. Such irregularities form the weak spots, where the metal will break first. Armourers, who aimed to make swords that were maximally resistant to violent use, learned through trial-and-error how to achieve a maximal hardness. The technique consists of heating the metal several times, each time letting it cool down slowly. High temperatures make the molecules move more forcefully relative to each other, and thus lead to variation. This variation breaks weak connections and gives the molecules the chance to discover a more stable configuration. During the cooling down, the variation gradually decreases, so that the metal can settle in the most regular, and thus the most stable, structure.

- Experiment: self-organisation of paperclips



Fill a box with paperclips that are opened slightly, so that they can slide into each other. Now shake the box. Result: the paperclips self-organise into forked chains. Explanation: it is easier for two loose paperclips to slide into each other, and thus get connected, than to get disconnected again. The selection principle is therefore: paperclips that are slid into each other are "preferred". The harder and longer you shake the box (more variation), the more paperclips will be hanging together in the end.

- Blind variation

In the theory of evolution, we often speak of "**blind**" or "random" variation. Mutations in genes are in general purely accidental and unpredictable. The essence of the Darwinian theory of evolution is that this suffices for evolution, on the condition that there is also selection. This means that the variations do not need to be directed to still produce directed evolution. The variations do not have "foreknowledge" about the right direction of evolution; they cannot "foresee" their further trajectory; they are "blind". The examples that we have discussed show that we indeed do not need anything else for directed evolution.

However, this does not rule out that variations *can* be directed, as is the case in many situations. The fundamental insight is however that evolution would work *even* if variations were 100% blind. This allows us to explain organisation without needing to appeal to any existing order or intelligence ("God") that directs the process. (We still need to explain where exactly this selective pressure comes from, which we will do by introducing the concept of "fitness".)

Another way to formulate "blindness" is that variation does not "know" what selection's preference is: variation and selection mechanisms are in general independent of each other. For example, mutations in DNA are independent of the environment that the organism is trying to adapt to. The movements of the carpet for instance are independent of the direction of the hairs on the carpet.

- Evolution as problem solving

A universal method to solve problems is the following: try something (trial) and if it does not work (error), try something else, until you eventually reach your goal. This is the method that evolution uses to search for "preferred" systems.

If you have solved similar problems before, you will however in general not try something purely at random, but start with things that you know, from experience, to have a larger chance of success. In this case, your search is not really "blind", but on the other hand, you do not know exactly how to reach your goal either. This is called *heuristic* problem solving: although there is no guarantee of success, you do have a more or less efficient method or intuition that allows you to reach your goal a bit faster. As we will see below, many systems have in the course of evolution developed forms of such "heuristic" knowledge, so they no longer need to grope around blindly. However, they still do not have a clear vision for the future. In this case, the general idea of (non-blind) variation and selection remains useable to describe the process, rather than a vision based on predestination or determinism.

10.6 The stepping stone principle

We have seen that more variation increases the chances of finding a solution to a problem. However, this is of little help if the chance of success is *a priori* immeasurably small. Typical organisms are after all immensely complex systems that can only survive in very specific circumstances. "Very complex" means many components and properties and therefore a very large state space. "Very specific circumstances" means that only a very small part of the state space is viable. The percentage of viable states is therefore infinitesimally small. The question that now arises is: how has blind variation managed to discover exactly these specific states?

- Example: evolving a very simple organism

Consider an imaginary organism with a DNA chain that is 1000 "words" long. (This is in fact a lot simpler than even the simplest organisms we know.) Every "word" represents one amino acid (the building blocks of the proteins that make up the cell). There are 20 amino acids that are used by life on earth. These form the "vocabulary" of the genes. The state space for this system consists of all possible "sentences" with a length of 1000 words that can be made, where each word is chosen among the 20 possibilities. The number of possible states is then 20^{1000} or about 10^{1301} (1 followed by 1301 zeroes)!

Imagine that only one state is the right one. How long will it take before you can find that state through variation and selection? Assume that you would make one guess every second, and that you would start guessing at the beginning of the universe (about 10 billion years ago), then you would have had less than 10^{17} chances to find the right sentence, which is only a minimal fraction of what you need. Guessing faster (for example a thousand times per second) does not make it less hopeless. Conclusion: no matter how much variation there is, you will never find the correct solution by only guessing blindly.

- Example: Hoyle's Boeing

Imagine a tornado passing through a junkyard. How big is the probability that blind variation will assemble various pieces of scrap metal and plastic in the shape of a Boeing 747? Again the probability is virtually zero, independent of duration or the power of the tornado.

Such examples are typically used by people who are sceptical about Darwin's theory of evolution. (This example comes from astrophysicist Fred Hoyle, who believed that life could not have arisen on earth, but needs to have an extraterrestrial origin.) According to them, the chances of assembling a living being through blind variation is so small that we have to postulate another type of mechanism to explain life, such as an intervention by God, by extraterrestrials, or by a kind of power (self-organisation, *élan vital*, ...) that was not foreseen by Darwin. The divine explanation has two versions:

1. *Creationism*: God has directly created all living beings, as described in the Bible. Problem: how do you explain fossils of primitive animals? And who created God?
2. "*Intelligent design*": an unspecified intelligence has now and then intervened in evolution in order to steer it in the right direction. This is more difficult to refute, but still does not give a real explanation: where does this intelligence come from? How does it work? Why is it intervening?...

The following **stepping stone principle** however gives a much simpler explanation.

- Example: the parable of the safecracker

The following example was proposed by H.A. Simon. Imagine a safe with a combination lock consisting of six digits. To open it, you need to find the correct combination of digits. Six digits with each ten possible values gives $10^6 = 1$ million possible states. This is much smaller than the state space of an organism, but in practice still far too large to find the solution by guessing.

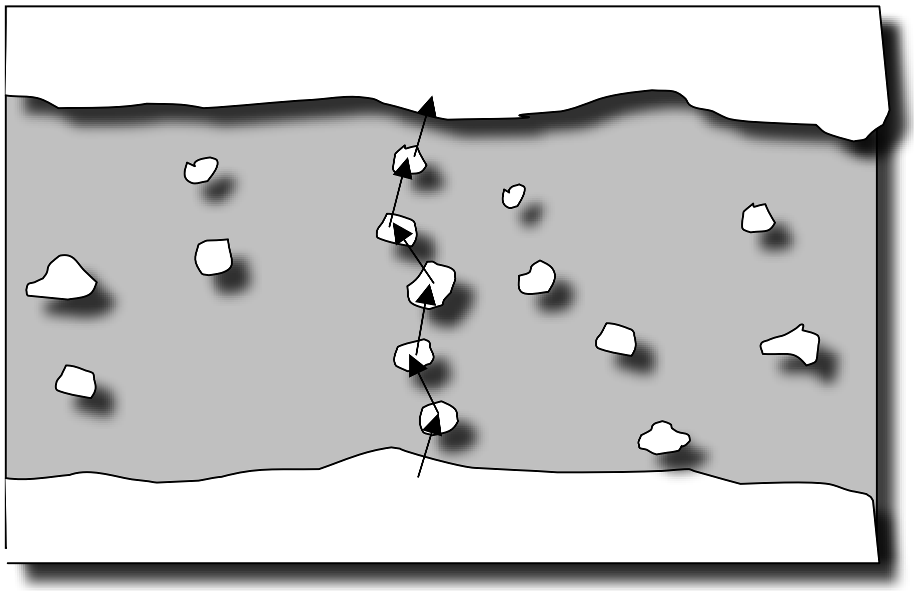
How can a safecracker open the lock? Each of the cogs of the combination lock has 10 positions. One of these positions, the right one, has been used often and is lightly worn. When the cog goes through that position, you hear a light "click" (this is the reason that safecrackers listen carefully with a stethoscope). You now know that this cog is in the correct position, and that you can move on to the next one. After maximum 6×10 trials all cogs are in the right position.

Instead of having to try 1,000,000 times, you got there with not more than 60 trials! The crucial step is hearing the "click". This indicates that you are on the right path and that you can now turn your attention to the next cog. You no longer have to try all possible combinations, because you can leave the cogs that you have already tested in place.

In essence: the "click", that is, the information that you are on the right path, that you have a part of the solution, reduces the number of possibilities to try from a multiplication ($10 \times 10 \times 10 \times ...$) to a summation ($10 + 10 + 10 +...$). The number of possibilities no longer increases exponentially, but linearly.

<u>Example:</u> now let us apply the same assumption to the imaginary organism from the previous example: $20 + 20 + 20 \ ... = 20 \times 1000 = 20{,}000$, instead of 20^{1000}. With 20,000 tries you are certain to find the solution. For evolution, this is trivial. There have after all been billions of organisms over billions of years to try the different possibilities.

Now consider a more realistic organism with 1 million words in its DNA. Then we need 20×1 million = 20 million tries. This is still very little by evolutionary standards.



In practice we of course do not get a “click” with every word that we guessed correctly. However what we do find are partial solutions: “intermediate steps” to the more remote goal.

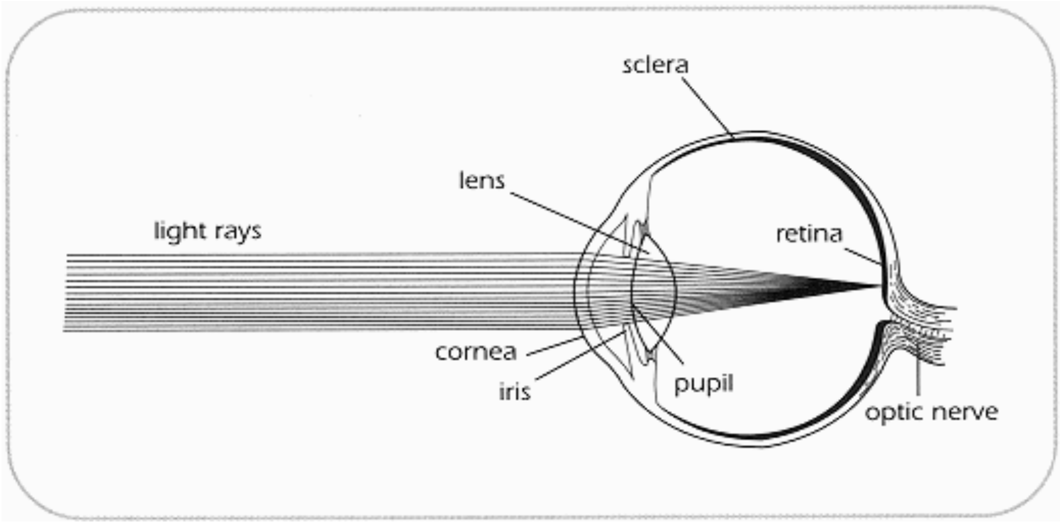
- The metaphor of the river

Imagine that a river is 6 meter wide—which means that it is too wide to cross without getting wet. Now imagine that after every meter, there is a big boulder sticking out of the water. You can then step from stone to stone until you have crossed the river. This is how partial solutions function as “stepping stones” towards the final solution and make a seemingly unsolvable problem almost trivial. (Note that there is no such thing as a “complete” or final solution in evolution: there are only partial solutions, and everything can always be improved.)

In the illustration above, a state space has been represented two dimensionally. White surfaces (“stones”) represent states that are viable or fit; grey surfaces (“water”) represent states that are not fit, and where the system would be eliminated. Variations can jump over grey surfaces on the condition that the jump is not too long. It suffices that there are enough white surfaces (“stepping stones), that are not too far from each other, to allow evolution to cross the large grey zone (“river” via some small jumps (the arrows in the illustration).

<u>Example*:</u> the evolution of the eye in six “steps”

The complex organisation of the eye is another example that critics of the theory of evolution like to use. If you blindly throw together some components, chances that you will get a working eye are tiny. Neither are there, at first sight, partial solutions: what can you do with a quarter or a half eye? Yet it is actually rather simple to imagine a sequence of intermediate steps in the evolution of the eye:



1. A part of the skin becomes light sensitive: the animal can now distinguish day from night.
2. The light-sensitive part gradually sinks away into a cup that forms a simple "pinhole camera", so that the light through the opening ("pupil") is projected onto the light-sensitive part ("retina") and thus forms a blurry image. The animal can now distinguish from which direction the light is coming.
3. Transparent skin grows over the cup and closes it; the light-sensitive tissue is now protected against dust, heat and cold.
4. The transparent skin thickens in the centre: this forms a simple lens that focuses the light, so that the image becomes sharper.
5. Muscles pull the lens to make it thicker or thinner: the eye can now focus on both nearby and far-away objects.
6. The light-sensitive cells mutate into different types, sensitive for blue, green and red light, respectively: the eye can now distinguish colours.

Each of these steps is a clear improvement over the previous step, and will thus be selected for. This evolutionary sequence seems realistic, since there exist animals exhibiting each of these stages of development.

10.7 Exaptation and the paradox of irreducible complexity

Despite all of these arguments in favour of evolution, the opponents of the theory of evolution have not surrendered yet, and they continue searching for some or other form of higher intelligence that directs evolution. Thus far, the most sophisticated argument from the "intelligent design" camp has been articulated by biochemist Michael Behe. He defines a system as "irreducibly complex" if leaving out only a single component means that the system no longer functions. This implies that the function of the system emerges from the *whole* of all of its components, and thus cannot be reduced to a combination of the functions of the individual parts. Besides a couple of complicated examples of biochemical reactions in living organisms, Behe illustrates this with the example of a mousetrap. An even simpler example of an "irreducibly complex system" is a clothes peg. It consists of two wooden slats that are pressed together with a metal spring. It is clear that the clothes peg would be completely useless if one of the components would disappear: without the spring, the slats fall apart, and if we take away one slat, there is nothing to press the other slat against.

According to the stepping stone principle, a complex system will evolve via a number of intermediary steps. Behe assumes that each step produces one component of the system, since it is unlikely that blind variation would create more than one component at a time. With an irreducibly complex system, however, all components must be present together for the system to function. The intermediary steps therefore provide no benefit whatsoever: they only produce components that are as yet still useless. Natural selection will therefore not retain them. According to Behe, natural selection in this case can only work if all components would all at once end up in their right place. As illustrated by Hoyle's Boeing, the probability that this would happen through blind variation is vanishingly small. Therefore we have to assume an intelligent force that has brought the components together in the correct way.

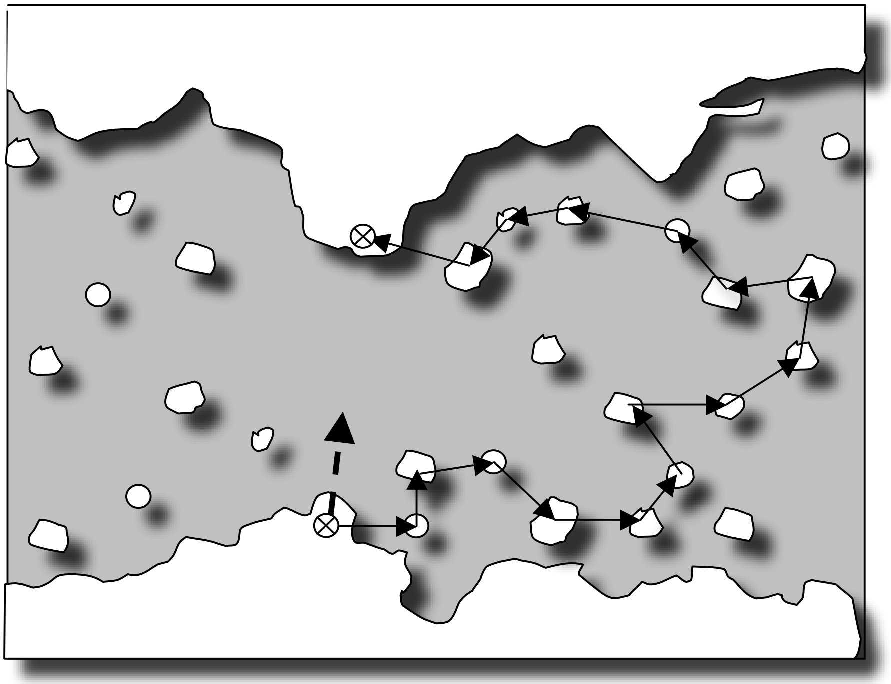
The mistake Behe makes in this reasoning is to assume that the function of a component as we know it now would also have been its function when it first appeared. This would imply that evolution works in a directed manner and adds the components one by one with a view to their future function—just like a carpenter screws a leg to a board with the intention that this, after adding another three legs, would eventually form a table. When there is only one leg, it cannot fulfil its function of supporting the tabletop, since at least three legs are required to form a stable whole. Variation is however a "random walk", now going in one direction, then in another, pausing only when it meets a stable, "fit" configuration. Therefore, in general these intermediary steps form anything but a straight line from start to finish.

In many cases, the intermediary steps will be fit because they fill a completely different function than the one performed by the system where the process finally ends up. While evolution proceeds, components formed thus will regularly *change function,* if it turns out that they also have another use. This is called **exaptation**: a component that was originally adapted to fill a certain function (adaptation), later appears to be useful to fill a completely different function—making the original function secondary or maybe even useless. Evolution is *opportunistic*: it uses the possibilities that present themselves at that moment, without looking ahead or planning, or trying to find the shortest way to a distant goal. Let me illustrate this with some examples.

- Examples: swim bladders and feathers

The swim bladder that allows fish to regulate their vertical position under water was originally a primitive lung that allowed lungfish to gasp for air above the surface when there was not enough oxygen in the water. Most modern fish live in well-oxygenated water and therefore no longer need to breathe air. Their lung function has disappeared. However, their descent from lungfish has given them a clear advantage compared to more primitive fish, such as sharks and rays, which do not have a swim bladder and therefore have to keep moving to prevent sinking to the bottom.

We find a similar form of exaptation in what appears at first to be an example of irreducible complexity: birds' feathers. Birds can fly thanks to the remarkable structure of their feathers, which allow air to pass through when the wing moves upwards during a wing beat, but not when it moves downwards. It is generally accepted that birds have evolved from small dinosaurs that had neither wings nor feathers. To learn to fly like birds, these dinosaurs needed to develop both wings and feathers, since either component alone does not suffice. (Note that there were flying reptiles without feathers, but these flew in a completely different way—or rather they glided.) We can however solve this problem quite simply by noting that feathers have another function besides flying: they are highly suitable to maintain body heat, as anyone who has slept under an eiderdown duvet can attest. It is therefore likely that feathers first arose as protection against the



cold, and only much later, after the ancestors of birds had learned to glide from tree to tree and thus had developed wing-like limbs, acquired their specific flight function.

- The importance of detours

These examples are very simple, because they do not comprise more than two components or functions, but the same reasoning applies to more complex systems with dozens of components. The examples of “irreducible complexity” that Behe put forward in his book have all been refuted using evolutionary “roadmaps” that use **exaptation**.

We need to add that evolution does not always make a system more complex, but sometimes streamlines and simplifies a complicated design, because this makes it more efficient to build and use. Imagine that a complicated series of intermediary steps leads to the evolution of a system with 20 components, which each offered a useful contribution at a certain point in time. Now that these components are working together perfectly for their collective function, it turns out that this function can also be performed with less elements. Selection for greater efficiency will in general lead to the gradual disappearance of components that are no longer useful, until each remaining component is absolutely necessary. A follower of Behe might conclude that the system shows irreducible complexity, and then decide that it could not have developed via intermediary steps.

This is understandable if you reason like an engineer or designer who has a clear goal in mind, and tries to conceive the shortest or most logical way to that goal. But what we, with hindsight, would consider the most logical way, is independent of the path evolution actually followed. The state space for a complex evolving system is so immense that we cannot imagine even a fraction of the possible trajectories that lead from state A to state B. Our natural tendency is to consider only the shortest or most direct paths. If we do not find enough stepping stones on this path, we will tend to conclude that evolution would never have been able to bridge the distance on its own. However, blind variation is not subject to these biases and will therefore not hesitate to follow the most complicated twists and turns, exploring half a dozen eventually unnecessary functions on the way, to eventually end up in B, albeit from a completely different direction than we would have expected.

Chapter 11. Fitness

11.1 Recapitulation: variation and selection

Variation is the exploration of new states, without bias or direction. Because of variation, the possible trajectories through the state space will diverge or fan out, resulting in a growing variety of possible states (distinction creation). As a result, variation increases variety or entropy, and decreases the information we have about the system.

Variation does not need to be explained. We can see it as a consequence of random, uncoordinated influences or disturbances. As long as the universe is not completely rigid or frozen, there are always sources of variation. Every system of which the temperature is above absolute zero (–273°C), undergoes thermal fluctuations (small, random movements of molecules). According to quantum mechanics, even the vacuum (which is by definition at absolute zero) will undergo quantum fluctuations. Chaos theory adds that such microscopic fluctuations can produce macroscopic effects (the butterfly effect). Since a system interacts by definition with the rest of the universe, it will necessarily undergo the influence of such accidental changes, and therefore spontaneously change state.

Nevertheless, to get a constructive evolution, variation has to be complemented by selection. **Selection** is the preferential retainment of certain states or variations, and the elimination of others. In this way, selection is the direct counterpart of variation, which reduces the number of states and therefore the variety or entropy. Selection is that which restricts a diffuse "cloud" of states from diverging too much, or makes it converge to a particular region of the state space, i.e. an attractor. Accordingly, selection reduces our uncertainty and increases information.

But what exactly is being selected? The question is why certain states are preferred above others. Where lies the difference? Which are the "forces" or the dynamics that propel the system in a certain direction? The answer is **fitness**: states with a higher fitness are preferred. The problem now is that we need to understand exactly what fitness is, where it comes from, and how it directs the evolutionary process.

11.2 The tautology of natural selection

Natural selection can be defined as the "survival of the fittest": the fittest are retained, the less fit are eliminated. This formulation can however be seen as a *tautology*, that is, a statement that is necessarily, by definition, true. Indeed the most general definition of "**fit**" is "that which survives". (The more traditional interpretation of "survival of the fittest" as "survival of the strongest" appears to be incorrect, because "strength" has little to do with evolutionary success.) According to this interpretation, natural selection means nothing more than *what survives, survives*.

Such a statement appears trivial or meaningless, and various thinkers have rejected natural selection as a general explanatory model for that reason. The interpretation of natural selection as a tautology is not strictly speaking incorrect, but the conclusion that the principle of natural selection is therefore useless, misses the point. After all, the laws of logic and mathematics are all tautologies. That does not mean that logic and mathematics are trivial or meaningless: the propositions that can be proven on the basis of these axioms are often very complex and counterintuitive, and their applications for calculations or deductions are legion.

<u>Examples:</u>

- *If A is true, and if A implies B, then B is true.* This is the most fundamental inference rule in logic, the so-called *modus ponendo ponens*. The rule appears self-evident, since implication is defined by the property that you can deduce the truth of the second clause (B) from the truth of the first (A). The *modus ponendo ponens* is still a useful and fundamental rule, which allows both students and computer programs to make the correct inferences.
- The axiom of contradiction: A *&* *not* A = *false*.
- The foundation of addition: 1 + 1 = 2.

The use of such implications or equalities that are at first sight trivial is the connection between two different representations, viewpoints or aspects of the same phenomenon (for example 1 + 1 and 2). If you know the one representation, then you can deduce the other. Changing the representation often allows you to infer non-trivial predictions or insights.

We can apply the same reasoning to natural selection: that which is selected and that which is fit are two perspectives on the same phenomenon. "Fit" considers the properties of the system in relation to its environment. "Selection" considers what happens with the system in the long run. The rule "that which is fit will be selected for" (and vice versa) helps us to make the correct deductions, and to come to conclusions that are in general not at all self-evident. In the example of the giraffes and their long neck, we will immediately start to imagine their specific environment, a savannah with tall trees, when asked "what is it that makes a giraffe fit?" From the general knowledge that an animal needs to find enough food to eat in order to survive, we can then deduce that fitness in this specific case requires a long neck. Thus, the principle of natural selection is in the first place a heuristic, meaning a rule of thought that allows us to divide complex problems into more comprehensible subproblems.

The advantage of a tautology is that such a statement does not have to be justified: it is by definition true. If you have reduced a phenomenon to natural selection, then you have in a sense *completely explained* it. On the other hand, when you reduce a phenomenon to the laws of nature or to the will of God, then you will still need to explain why God would have wanted this, or why the natural laws are like this.

11.3 Definition of fitness

The intuition that we want to express is as follows: fit systems or states become more numerous; unfit systems become less numerous and eventually disappear. We can quantify this as follows:

This means that the fitness F of the state / system s is equal to the number of appearances N of s at time t + 1 (future generation), divided by the number N at time t (current generation). This is the definition used in genetics. We can now distinguish the following special cases:

- $F(s) = 1$ means that the number of appearances of N remains constant.
- $F > 1$ means that the number increases.

- $F < 1$ means that the number decreases, and that the systems of type s will eventually die out.

There are different mechanisms that determine whether the number of "appearances" increases or decreases:

- **Survival**: if not all appearances survive, the number decreases, which leads to a low fitness.
- **Reproduction**: if each system produces a lot of offspring, the number increases, which leads to high fitness.
- **"Spontaneous generation"**: if the systems of a certain type arise spontaneously (i.e. not through reproduction of an existing system), the number increases, which leads to high fitness.

Living organisms do *not* arise spontaneously, but certain physical systems (e.g. molecules, snowflakes, crystals etc.) do. Some of these (such as autocatalytic molecules, see 14.5, or crystals) can either increase through reproduction of the base form (for example a salt crystal that is dropped in a saturated saline solution) or through spontaneous generation (even without the added crystal, salt crystals will eventually form in the solution). Here we have to note that since most applications of fitness and evolution come from biology, spontaneous generation is usually overlooked as a fitness mechanism.

To conclude: fit appearances do not die off too quickly, and are either reproduced quickly enough or arise spontaneously.

There are different "strategies" to obtain high fitness, however, depending on which of the three mechanisms (survival, reproduction, spontaneous generation) plays a more or less important roles. Indeed, it is in general not possible for a system to excel in all three properties. For example, systems such as organisms are too complex to be generated spontaneously, through self-organisation. Because of their complex organisation, they can however maintain a subsystem that is specialised in efficient reproduction: the reproductive organs. We will now generalise this idea of different strategies.

11.4 Fitness dimensions

Since every system and every environment are unique, fitness has in principle an infinite number of different aspects or dimensions. The properties that increase fitness for one type of system (e.g. a long neck for giraffes) are in general irrelevant for another type (e.g. viruses), where completely different properties are crucial (e.g. the ability to avoid the immune system of the host). Yet with our abstract definition of fitness, we can still distinguish certain universal "dimensions" according to which fitness can vary.

- r-K selection

Two general strategies to obtain fitness can be distinguished for organisms. (The labels r and K come from a classic mathematical model of population growth.)

r-selection is characterised by fast reproduction and growth, but a short lifetime. Examples of organisms that follow this strategy are bacteria, insects, mice and weeds. These organisms are typically small, vulnerable and not very sophisticated. r-selection takes place in an environment where there is plenty of food to grow fast and reproduce fast, but where life is dangerous and unpredictable (for example because of predators, illnesses or strong fluctuations in the amount of available food), so that it is pointless to aim for the long term. Since survival of a certain individual will in general depend on uncontrollable, accidental factors, it is best to produce as many descendants as possible in the hope that at least one will survive. Natural selection here prefers quantity over quality: it is pointless to invest a lot of energy in individual descendants, by for example making them big and strong or by having them go through a long learning process, because the large risks in the environment mean that there is no guarantee that these extra trumps will improve the chances of survival.

K-selection is the other extreme of the continuum reproduction ↔ survival. This strategy aims for a long lifespan, but the downside is slow reproduction and growth. Examples are humans, tortoises, elephants, oaks and other trees with hard, durable wood. Such organisms are typically large, well protected and because of their long lives, they are able to gain a lot of experience. Their environment is relatively safe or stable, but because of mutual competition, the amount of food and resources is limited, so that there is no point in producing many descendants. Selection here prefers quality to quantity, and the parents invest a lot of energy to make their individual offspring as fit as possible so that they can handle the competition with others as well as possible.

The differences between r- and K-selection are not only found between different species, but also within a species, when it is useful for an individual to choose the strategy that is best adapted to the individual situation. An example that I noticed while observing growth in my garden is the following. Some plants reproduce vegetatively through their rhizomes (networks or roots), which form long offshoots below ground, which then suddenly shoot up a meter or more away from the parent plant and form a new stem. I had noticed that these new stems appear much earlier in spring and grow much faster than the parent plant. At first sight, this is paradoxical, because you would expect that most of the food reserves are still in the old rhizome. According to r-K logic, on the other hand, this is an efficient strategy: the old rhizome is in a relatively safe, predictable environment, where the plant has after all already survived for years without problems. The new offshoots on the other hand are in unknown terrain, where there might be a shortage of light, food or water, and a danger of being pushed away or overshadowed by other plants. The offshoots therefore need to conquer the terrain as fast as possible, even if they have to run the risk of developing leaves early and in that way expose them to late frost (r-strategy). The parent plant on the other hand, has little to fear from being pushed away and can therefore afford to grow more slowly (K-strategy). This reduces the risk of freezing and helps to form qualitatively strong stems and leaves.

A more interesting (and more controversial) application of the r-K logic is human reproductive behaviour. It is an iron law of demography that the rich, developed countries have a smaller birth rate *and* death rate than poorer Third-World countries. Moreover, both figures fall dramatically when the country develops—this is called the *demographic transition*. The lower death rate is easy to explain: in more developed countries, there is more investment in services such as food production, healthcare and safety, so that the risk of common causes of death such as famine, illness, accidents or war becomes a lot lower. The lower birth rate on the other hand is a lot less obvious. It is a paradox that exactly in the poorest countries, where there is hardly enough food to keep people alive, seven or eight children per woman are standard, while in countries where there are plenty of facilities to support child care, the birth rate is between 1 and 2 children per woman. Moreover, within those rich countries, it turns out to be the underprivileged or oppressed populations that have the most children. Examples are African Americans and Hispanics in the US, Islamic immigrants in Western Europe, the Roma in Eastern Europe, and the Palestinians in Israel.

If we however assume that people in difficult circumstances (subconsciously) follow an r-strategy, then it follows that exactly these populations would have the most children. Other typical properties of underprivileged and poor groups also follow the r-pattern: babies have a lower birth weight; women become pregnant at a younger age; adolescents and adults tend more towards risky behavior such as smoking, drug use, unsafe sex, reckless driving, participation in crime or war..., in the hope of short-term advantages. However, they tend less towards long-term planing or investments, such as a studying for a university degree. Finally, their life expectancy is lower.

Conversely, people who have grown up in a safe, reliable environment will subconsciously follow a K-strategy. This implies that they will bring few children into the world, but in the long term invest a lot in the health, wellbeing and upbringing of these rare descendants. Moreover, they strongly tend to avoid risks.

The evolutionary anthropologist Chisholm has argued that these subconscious strategies are regulated by hormones: the unsafe environment in which underprivileged children grow up would lead to a higher production of the stress hormone cortisol. In turn, this would stimulate the production of the sex hormones estrogen and testosterone. These will then lead to the high fertility of women and the macho behaviour of young men that are typical for underprivileged populations.

- The two aspects of fitness

The word "fit" in English has two meanings, each of which summarises a complementary aspect of the selection process:

1. *Adapted*, *fitting*, *suitable*: in evolutionary terms, we can interpret this as system fitting into its environment and making optimal use of it. It survives external selection, i.e. selection by the environment (see 12.4). This is the external or relative aspect of fitness.
2. *Robust*, *healthy*, *in good condition*: this means that the system is able to survive autonomously, independent of the environment. It survives internal selection (see 12.4). This is the internal or absolute aspect of fitness.

- External or relative fitness

<u>Examples:</u> a piece of a jigsaw "fits" in the right slot. Once in place, it is hard to get it out again. A key "fits" in its lock. The molecule of a medicine fits in the correct receptor in our body.

This type of fitness is relative to the environment, and more specifically relative to the specific position that the system has within its environment, or the role that it fills in that environment. Such a situation within a larger environment to which the system can adapt is called a **niche**. A niche is a way of life that exploits specific resources available in the environment to survive.

<u>Example:</u> Koalas only eat the leaves of the eucalypt tree. Without these trees, koalas would not survive. Other organisms, such as kangaroos, live in the same environment. They do not need eucalypt trees. Kangaroos fill a different niche than koalas.

<u>Example:</u> Within our economy, there is a niche for companies that are specialised in repairing car exhausts. Without cars, these businesses would not survive.

Aiming for relative fitness will in general lead to *specialisation*, so that the specific properties of the niche can be exploited as efficiently as possible. For example, car exhaust companies are more efficient in their specific niche than garages that repair cars in general.

- Internal or absolute fitness

The second type of fitness is absolute, independent of the environment, or of the niche that the system occupies within it. One way to attain absolute fitness is intrinsic stability or rigidity. For example, a diamond is very hard and therefore almost indestructible, regardless of the environment. Another way is intrinsic flexibility or adaptivity. For example, rats, humans and cockroaches adapt to the most diverse environments. This kind of fitness favours the "*generalists*", the jacks-of-all-trades. Generalists tend to do better in an environment that changes a lot, so that new adaptations are always necessary. Specialists tend to do better in an environment that is very stable, so that they can fully exploit it specific resources.

11.5 The direction of evolution

Although evolution is in general unpredictable, it will by definition prefer systems with higher fitness. *This is why in general the average fitness will increase.*

In the short term, in stable environments, this will lead to increased specialisation, and therefore to relative fitness. On the long term on the other hand, when the environment undergoes drastic changes (such as climate changes), the specialists are the first to die out, because they are no longer adapted, and the generalists will take over. For example, imagine that because of new environmental laws only purely electric cars would be allowed. Then the niche for exhaust specialists disappears, while a niche remains for general car technicians. Although such generalists will quickly produce diverse specialists to fill all the niches of the new environment (this process of divergence is called *adaptive radiation*), they in general maintain the essence of their adaptivity (see 14.1). In this way, their selection leads to increased absolute fitness.

We still have to clarify how it is possible that evolution is intrinsically chaotic and unpredictable, but nevertheless that it has a preferred direction. This is best explained using an analogy.

- The metaphor of the mountain

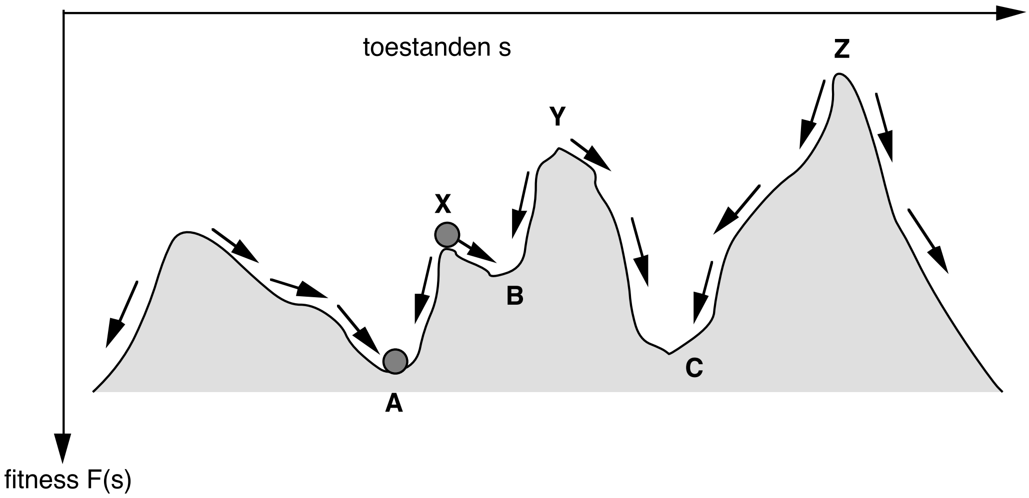
Let a ball go from the top of a steep, irregular mountain. The ball will roll down at high speed, crashing into rocks and other obstacles so that its trajectory is very unpredictable. You cannot tell where the ball will come to a halt. However, you do know almost certainly that the ball will stop at a position *lower* than the starting point. The direction of this evolution is unpredictable in the horizontal directions (east-west or north-south), but predictable in the vertical direction (down, not up).

We will now expand this metaphor into a precise, mathematical model, where fitness plays the part of the vertical dimension.

11.6 Fitness landscapes

Imagine the state space of a system as a two-dimensional, horizontal plane (in general a state space has many more dimensions, but this is hard to visualise). Suppose that for each state *s* we know the fitness value $F(s)$. Now imagine $F(s)$ in the vertical dimension, meaning that we are going to shift each point s vertically over a distance $F(s)$. So, $F(s)$ is the "elevation" of the point above (or below) the original plane. We have now transformed the state space into a **fitness landscape**. This means that instead of an even, uniform space, we now have an undulating landscape with hills and valleys.

Convention: To stick with the mountain metaphor, we will see *higher* as *less fit* rather than as more fit. This is merely a convention, which has no effect on the effectiveness of the representation. Note that in biology and in the CAS approach the traditional convention is the other way around: here a "fitness peak" corresponds to high fitness. This convention is actually a little less intuitive for understanding evolution in a fitness landscape, because the system has to actively "climb" to the top, rather than spontaneously roll downhill. On the other hand, in physics (where they speak of a *potential function* rather than a fitness function), the same convention is used as in the mountain metaphor, meaning that the preferred direction for evolving systems is *downhill*.



In the illustration above, the state space is represented as one-dimensional (x-axis, horizontal), because that is easier to draw. The fitness is represented in the vertical dimension (y-axis). According to our convention, the mountain peaks (X, Y, Z) correspond to low fitness; the valleys (A, B, C) correspond to high fitness.

The system by definition always prefers a fitter state to a less fit state, and therefore moves from the latter to the former. You can imagine the resulting evolutionary process as a ball that rolls down the mountain (for example X) into the valley (for example A) and stops there. The steeper the hill, the faster the ball (the system) goes down. The arrows in the illustration show in which direction the evolution will go for the different regions of the state space—for example from X to A or B, from Y to B or C.

From each state, the system will move towards the neighbouring state with the highest fitness. However, it cannot *jump over* the neighbouring states to a state further away, even if this is fitter. The reason is that variation is in general small: only one or a few properties are varied at the same time. A ball in B cannot roll over X to A, even though A is lower. This has important implications:

11.7 Local and global maxima

Every bottom of a valley (for example *B*) is a **local maximum** of the fitness function. This means that no state in the immediate neighbourhood of *B* has a greater fitness than *B* (in other words lies lower). A system that comes near *B* will necessarily have to end up in *B*. Once in *B*, it can neither move forward, nor back, because every neighbouring state now has a lower fitness and is therefore less good than *B*. The ball can only roll down, not up. The system cannot reach *A* or *C* from B, even if these states have greater fitness.

The **global maximum** is the state with the highest fitness of all. In general it is however very difficult for a system to reach this global maximum. The system can only try to find the fittest state *locally*, in its immediate neighbourhood. Variation is after all **blind**: it cannot anticipate the peaks and valleys further away and can only "grope" to find its way. There is a high probability that the system will become stuck in a local maximum that in fact has a very poor fitness, simply because there is no better alternative nearby. Even if a state with very good fitness were "just around the corner", the system would still not know how to find this maximum.

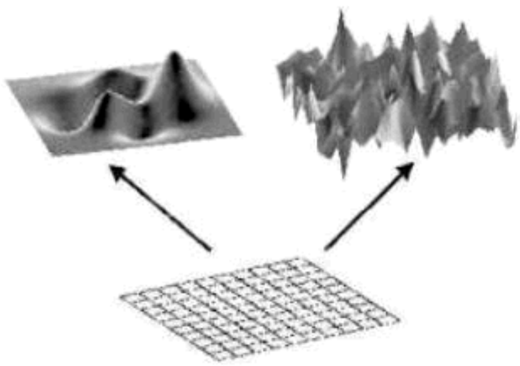
<u>Conclusion:</u> evolution does not optimise (find the best solution); it only tries to improve the local state until it can go no further.

This representation appears to imply that evolution stops as soon as the system has reached a local maximum, and that no further improvements are possible after that. Yet in practice we see that systems do not remain forever at the bottom of the valley, but sooner or later manage to make a big step forward. How does a system get out of the valley? There are two possible mechanisms:

1. The "order from chaos" principle: increase of variation enables bigger "jumps" and thus possibly a leap over a small mountain peak. In general, variations (e.g. mutations) are small, so that they remain in the immediate vicinity. If there are no fitter states in this vicinity, the system will always return to the local maximum. However, now and then a variation occurs (for example a "macro mutation"), that possibly ends up in a deeper valley, for example a leap from *B*, over *X* to valley *A*. Nevertheless, his happens rarely, because the risk of a drastic reduction of fitness is much greater for large leaps than it is for small leaps.
2. Change in the structure of the fitness landscape: mountains can subside or rise from valleys as a result of changes in the environment. States that used to have a low fitness can now get a higher fitness, and vice versa.

<u>Example:</u> a polar bear that is a little darker than its fellow polar bears stands out on the ice and will thus be less successful at catching seals. When the climate changes and the ice disappears, a less white pelt is no longer a disadvantage, but an advantage, and all bears will quickly evolve towards the darker colour that is adapted to their new environment. The white state is no longer at the bottom of the fitness valley, but has, through climate change, ended up on the top of a mountain.

The structure of the fitness landscape explains why the speed of evolution is often very irregular: sometimes very fast, sometimes standing almost still. Such a pattern of change is called "punctuated equilibrium": most of the time there is equilibrium and nothing much happens, but at certain times, this equilibrium is "punctuated", or interrupted, through sudden, catastrophic changes that produce new forms or species. A position in the middle of a broad, deep valley (such as *C*) halts evolution, since every normal variation entails deterioration (decrease of fitness) and is therefore selected against. A position at the edge of a deep valley (for example *X*, close to *A*) can however result in a sudden, fast descent into the valley.

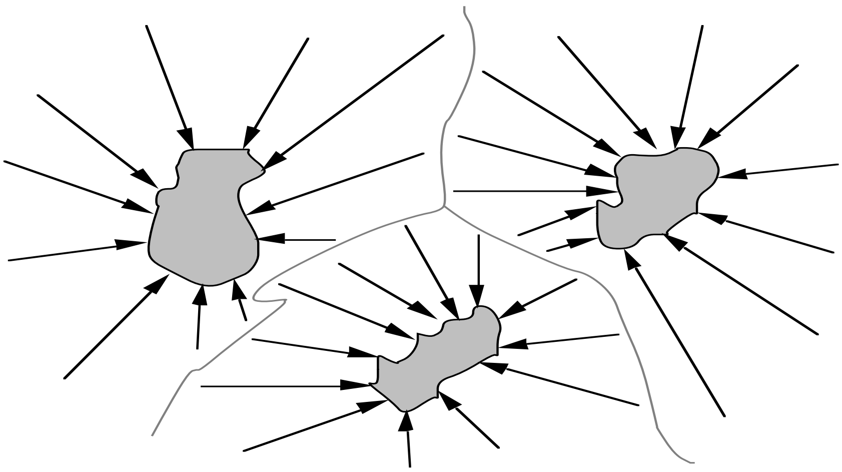


An important aspect is the degree of *ruggedness* of the landscape. A landscape can be gently rolling with only a few peaks and valleys, or it can be craggy with a large number of higher peaks and deeper ravines. The illustration shows how the same two-dimensional state space (below) can give rise to either a smooth (left) or a rugged (right) fitness landscape. The more rugged the landscape, the more irregular and unpredictable the evolution, and the harder it is to find the global maximum.

11.8 Attractors and basins*

A fitness valley can be seen as an **attractor** in the state space: a region that the system can enter easily, but out of which it cannot get on its own. The hillsides that surround the deepest point form the **basin** of the attractor: these are all the points in the state space from where further evolution will normally end up in the valley or attractor. The different basins are separated from each other through "mountain ridges": peripheries from which you will either go down into the one valley, or into the other. For example, in the one-dimensional drawing: *X* and *Y* are the mountain ridges separating *A*, *B*, and *C*.

The illustration below, which can be seen as a fitness landscape viewed from above, or as a **phase portrait** of a dynamic system, illustrates the principle. The grey areas are valleys or attractors; the white areas around them are basins. The arrows show the direction of evolution, and the irregular dividing lines represent the "mountain ridges" dividing the basins. The name "basin" comes from the analogy with water reservoirs: the attractors can be seen as lakes in a mountainous landscape. The arrows then show the direction along which the rainwater flows down the hills, and the dividing lines show the divisions between the reservoirs or basins.



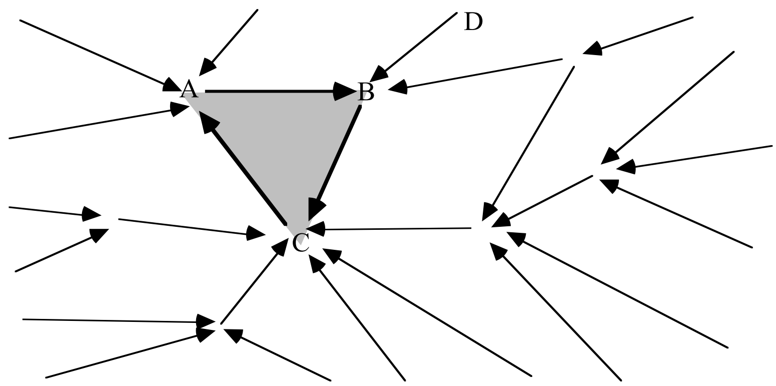
Note that the concepts of attractor and basin apply to a wider variety of phenomena than the concept of fitness landscape. Indeed, they can be defined even for processes where no fitness value can be determined. Attractors can also exhibit more complicated shapes or behaviours than the simple maxima of the fitness function. For example, a "limit cycle" is a closed loop in the state space along which the system moves continuously, without ever halting. The defining property of an attractor is however still that of a region in state space that the system can enter, but not leave. In that sense, an attractor is an essential illustration of the principle of asymmetrical transitions: certain transitions are preferred above others, even when you cannot attach a fixed fitness value to the resulting state.

Fitness, like all concepts in this book, is defined relationally, which means that it is always determined *relative to* something else. You can never talk about the objective, absolute fitness of a system or state, only about its fitness relative to another state and relative to a certain environment. Higher fitness of state A relative to that of state B means that selection prefers A to B, and that the chance that B will be replaced by A is larger than the chance that A will be replaced by B (what we called an "asymmetrical transition" in 10.4).

<u>Example:</u> a rabbit that can run at 30 kph in an environment where the foxes can only go 20 kph will be a lot fitter than a rabbit that can only go 15 kph. This means that there is a larger chance that this rabbit and its descendants will survive and take the place of the slower rabbits, than the other way around.

If we know these relative probabilities of survival, we can in general calculate an absolute probability of survival in the long term, and thus a fitness value. If we can determine a fitness value for every state, then we can represent this geometrically as a fitness landscape. In such a fitness landscape, an evolving system will always choose the direction "down", and end up at the bottom of a valley.

Now assume that we know the relative preferences of A over B, of B over C, etc., but that the following happens: B is preferred over A, C over B, but A over C. Assume that we could determine a fitness metric F for each of the states, then this would imply that $F(A) < F(B) < F(C) < F(A)$. This would imply that $F(A) < F(A)$: the fitness of A is both larger and smaller than itself! This is obviously impossible, and therefore we cannot determine an absolute fitness metric $F(A)$.



In this situation, A, B and C could still be part of an attractor, in the sense of a set of states that is preferable as a whole, even when no permanent preference can be made between the individual states within the attractor. The simplest case is a so-called limit cycle, where the system continuously passes through the transitions A → B → C → A → B → ..., and keeps returning to the same states. Although it is not possible to say that one of these states is fitter than the other two, it is possible that each of these states is absolutely fitter than a fourth state D, because A, B or C are always preferred over D, and the system never returns from A, B or C to D. Note that such a limit cycle by definition points towards a system far from equilibrium (see 3.6), because the system never stops changing.

Example: Consider the system consisting of rabbits and foxes together. If we no longer want to determine the fitness of an individual rabbit, but of the entire system, we will find that this is similar to a limit cycle, with the system regularly going through four states:

> (many rabbits, few foxes) → (many rabbits, many foxes) → (few rabbits, many foxes) → (few rabbits, few foxes) → (many rabbits, few foxes) → ...

The reason is that an increase in the rabbit population because of an abundance of food leads, after a delay, to an increase in the fox population, which in turn leads to a decrease in the rabbit population, because more rabbits are eaten by foxes. This is followed by a decrease in the fox population, because there are not enough rabbits left to feed all mouths. Thus, the system does not end up in an equilibrium, but continues moving cyclically from one state to the next. We can therefore not determine an absolute fitness for each of these four states. We can however imagine states that are less fit than these four states. For example, a state (very many rabbits, very many foxes) will quickly disappear, because the rabbits die off both because they are being eaten by foxes, and because there is too little grass. Such a state will therefore never occur in normal circumstances. This state belongs to the basin of the attractor formed by the limit cycle above: the trajectory through that state will quickly end up in one of the states of the limit cycle.

Chapter 12. The systems approach of evolution

12.1 Introduction

As we have noted before (2.4), the problem with traditional Darwinian or neo-Darwinian theory is that it is too reductionist, being only concerned with a single component (an organism or a gene) in a given environment. Self-organisation, systems theory and the CAS approach have drawn attention to other, more holistic phenomena that are also important in the evolution of complexity. Their emphasis lies on the manner in which different components interact and co-evolve, together forming a greater whole.

The question is how we can integrate all these approaches and phenomena in one coherent conceptual framework. The Austrian theoretician Rupert Riedl has suggested that we can solve this problem by developing a *systems theory of evolution*, which situates the concepts of variation and selection in a more holistic or systemic frame. The problem is that the more components and interactions we consider, the more complex the model becomes. This in turn makes it harder to produce predictions or explanations. The CAS approach has tackled this problem by relying on computer simulations, but through this, it has lost some of the conceptual clarity and intuitive understanding. Still, systems theory offers some very simple conceptual tools, such as the subsystem—supersystem distinction, that can be directly applied to the processes of variation and selection. As we will show now, this allows us to classify the different types of evolutionary processes in a simple manner, and integrate them in a large framework. In the next chapter, we will go in the opposite direction, and apply the evolutionary approach to understand where different types of systems come from.

12.2 Sequential vs. parallel variation and selection

We will begin by applying the distinction between sequential and parallel (section 4.4) to the elementary processes of evolution. This gives us the following types:

- Sequential

Sequential variation means that one state follows another one. There is a strict sequence of different states. At any given time, only one state is being tried out by the process.

<u>Example:</u> a bottle in the sea is propelled by wind and waves. The bottle can only be in one place at the time.

Sequential selection means that one of the states in the sequence is retained, so that the sequential variation stops. This means that all other states of the sequence are being "rejected" or eliminated.

<u>Example:</u> the bottle washes ashore on an island, and remains there. The stranded state is selected for, the drifting states are eliminated.

- Parallel

Parallel variation means that different states are being tried out in parallel, *at the same time*. This assumes that at a given moment different specimens or copies of the system exist, each having its own state. *Parallel selection* eliminates the maladapted specimens and retains the adapted ones.

<u>Example:</u> bacteria develop resistance to an antibiotic.

At any given time there are billions of bacteria present in the body of the patient, all descended from one or a few introduced germs. These bacteria continuously reproduce through division. Every new bacterium is potentially a variation on the other, because of mutations: small mistakes in the copying of DNA of the bacterium. All together, the bacteria exhibit an enormous number of variations on the basic pattern. Some of these variations are more resistant to the antibiotic and will therefore survive longer. If the bacteria manage to produce enough descendants with enough variations, they will sooner or later discover a state that is resistant and that will thus continue to reproduce.

- Relationship between both types

Parallel variation is the typical form of variation as conceived by Darwin and the biologists inspired by his theory. They assume that living organisms have multiple offspring that are all different. These descendants live side by side (in parallel), and in mutual competition. The larger the number of parallel variants (number of descendants, population), the larger the chance that one or more of them will have a state that is selected for. The more horses you bet on in a race, the larger the chance that one of them will win.

Thus, parallel variation is intrinsically more efficient than sequential variation. That does however not mean that sequential variation cannot lead to fast evolution. Sequential variation can be made more efficient by speeding up variation (i.e. going through a larger number of different states per unit of time). This is for example what we find in the "order from chaos" principle: the harder you shake the box with paperclips, the faster you reach a state where the paperclips are all hanging together.

Sequential variation is more typical for physical, computational or mental systems. Conscious thought is largely sequential: we can only focus on one potential solution at a time and if it turns out that it does not fit, we focus on the next variation, etc.

Evolution of knowledge happens both in parallel and in sequence:

- Sequentially: a scientist of philosopher considers different possible solutions for a problem, one after the other (variation) until one is found that solves the problem (selection).
- Parallel: different thinkers search for a solution to a problem simultaneously. If one of them is successful and finds (part of) the solution, the others will in general adopt this solution and combine it with their own results.

Most computers still work sequentially: they cannot execute more than one instruction at a time. It is however common to simulate parallel processes on computers. Your computer can for example run several programs at the same time. In fact, these programs do not run perfectly synchronously, but the processor spends a few milliseconds on the first, then the next, then the next, ... returning to the first, and so forth. This process goes so fast that it seems as if the different programs are executed at the same time. Because processors are so fast these days, they can simulate a "population" of systems varying in parallel, for example an artificial ecosystem with virtual organisms that evolve simultaneously.

Traditional evolution theorists, who work in a biologically inspired framework, in general do not consider sequential variation and selection. For them, it is still "replication" (making parallel copies with small variations) that is necessary for evolution. The fact that computers work sequentially and still are highly effective in the simulation of parallel evolution shows however that there is no fundamental distinction between both types.

12.3 Internal vs. external variation

Another fundamental distinction made in the systems approach, namely the distinction between what is inside the system boundary (internal, system) and what is outside of it (external, environment), also yields an interesting analysis of the different types of evolutionary processes.

Variation is in general *internal*, meaning that the components or properties of the system vary on their own, without any outside input. Examples are mutations in the DNA of an organism, and thought processes in the brain. Essentially this means that the system visits different states in its own, predefined state space. Most evolutionary theories assume internal variation, because this is easier to describe.

Nevertheless, variation can be external, meaning that there is an exchange of components with another system. In a sense, this implies the creation of a new system, with new components, and therefore a new state space.

Example: Sexual reproduction: pieces of DNA of the mother and the father are recombined to form the DNA of the child. Note that because the DNA of the father and the mother are highly similar (they belong to the same type of system), in practice you can describe recombination in the same way as mutation, i.e. as internal variation within a given state space.

Example: **symbiosis**: two organisms of a different type are interdependent, and through co-evolution they have adapted so much to each other that they form, in a sense, a new organism. For example:

- Lichen is a symbiosis of a type of alga and a type of fungus.
- Eukaryotes (complex cells, such as those in our bodies) contain mitochondria, energy-producing organelles that are descendants of free-living bacteria that have penetrated the cell.
- In our gut live bacteria that are necessary for our digestion.

Example: chemical reactions: two molecules come into contact with each other and exchange electrons and/or atoms, resulting in one, two, or more molecules of a new type.

Na (sodium, metal) + Cl (chlorine, gas) → NaCl (table salt)

12.4 Internal vs. external selection

In the traditional theory of evolution, it is assumed that selection is external, meaning determined by what is happening outside the system. States that are not adapted to their *environment* are eliminated. For example, a woolly mammoth in a warm climate or a white rabbit in a dark forest are poorly adapted and will therefore be selected against in favour of shorthaired or dark variants.

Selection can however also be internal: *intrinsically* unstable states disappear. Such states are eliminated independent of the environment. For example, an irregularly magnetised piece of iron in which some of magnets point in the direction opposite to the direction of the majority is unstable.

Internal selection is typical for **self-organisation**: the system itself determines which state "works" and which does not. For example, the state in which all internal magnets are aligned is preferred above other states. Internal selection can also be found in biology. For example, a nonviable embryo is spontaneously aborted before it comes into contact with the environment.

Internal selection is essential in mental evolution. Scientists trying to solve a problem will already reject the majority of possible solutions (variations) based on their own, internal selection criteria (for example incoherent, too complicated, contradicting known facts, etc.). Only when the idea has survived this internal selection, will it be tested in the outside world, for example through a scientific experiment. This is the phase of external selection.

In general, each variation will be selected internally before it meets the outside world and there undergoes external selection. This explains why many variations (especially the mental ones) appear intentional or guided. The original variations were blind, but most of these have been eliminated by internal selection, based on knowledge gathered earlier, so that only the best ones remain.

12.5 Generalised theory of evolution

The traditional, biological version of evolution assumes *parallelism*, *internal variation* and *external selection*. This is however a very restricted perspective: you also need to take into account sequential processes, external variation and internal selection. If you do that, many processes that do not fit into the strict Darwinian framework, can be described perfectly well through variation and selection: for example, self-organisation, symbiosis, thought processes, chemical reactions, etc.

The fact that most authors do not distinguish between these different classes of processes produces much needless controversy and confusion. For example, some describe self-organisation as based on selection, while others claim that self-organisation belongs to a completely different category that has nothing to do with selection. The concept "*natural* selection" (selection by nature, that is, by the environment) is often used to indicate classical Darwinian selection, but in fact this helps very little for the prevention of confusion, because self-organisation is also a "natural" process.

On a more fundamental level, we have to note that from a system theoretical perspective, there is *no strict distinction* between the different types: what is external for a subsystem, is in general internal for the supersystem that comprises this subsystem, and vice versa.

<u>Example:</u> an embryo that does not survive because it is unable to implant in the uterus is the victim of external selection: it was not adapted to its environment (the uterus). From the point of view of the mother (the supersystem), on the other hand, the failed pregnancy is an example of internal selection: the embryo was eliminated before it came into contact with the outside world.

Example: the input of the father's DNA via sperm is external variation for the mother (system), but internal variation for the species to which both father and mother belong (supersystem).

A set of subsystems that are evolving in parallel can formally also be seen as one supersystem that is evolving in sequence, and vice versa. The state of the supersystem is simply the Cartesian product of the states of the parallel evolving subsystems (see 8.5).

Example: You can describe the movements of two billiard balls on a table with two parallel trajectories, each in a two-dimensional state space (x coordinate = long side of the table, y coordinate = short side), or with one, sequential trajectory in a four-dimensional state space (x coordinate ball 1, y coordinate ball 1, x coordinate ball 2, y coordinate ball 2).

Internal or external, sequential or parallel: in the end this is merely a matter of perspective. In practice the perspectives are important, though, because they can make a description impossibly complicated or very simple.

Example: If you were to describe a colony of a trillion bacteria, each with a hundred genes that can vary, as one sequentially evolving supersystem, you would need a state space of a hundred trillion dimensions. It is much simpler to describe the system from the point of view of one bacterium that explores a hundred-dimensional state space, and to assume that the other bacteria are doing the same in parallel.

Overall conclusion: From a system-theoretical perspective, *all* evolutionary processes can be understood as a result of variation and selection. This includes physical, biological, mental and socio-cultural processes. This philosophy is called *selectionism*, (*universal*) *selection theory,* or *universal Darwinism*.

Chapter 13. Supersystem transitions

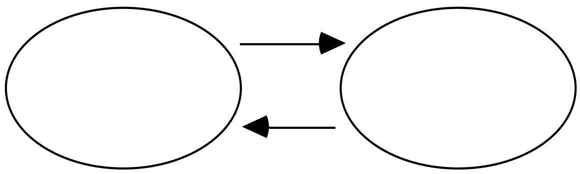
13.1 How does evolution lead to more complex systems?

In the beginning (just after the Big Bang that created the Universe), there were only elementary particles: no atoms, molecules or other, more complex systems. As evolution progressed, more and more complex systems were added. We will now research the basic mechanisms that have led to this complexification. In this chapter, we will focus on structural complexity, that is, the static structure of systems, rather than their dynamical behaviour or function. The question that we will discuss is, how do a number of components (subsystems) come together to form a supersystem? This evolutionary leap to a higher level of complexity is what we will call a **supersystem transition**.

13.2 Interactions

All systems or components are capable of *interactions* with other systems. If this were not the case, we would never be able to observe these systems. According to the principle of the identity of indiscernibles, this would mean in practice that they simply would not exist.

Interaction can be seen as action followed by reaction: the state of the one system influences the state of the other system; output of the one becomes input of the other. The state of the other system in turn in general influences the state of the one.



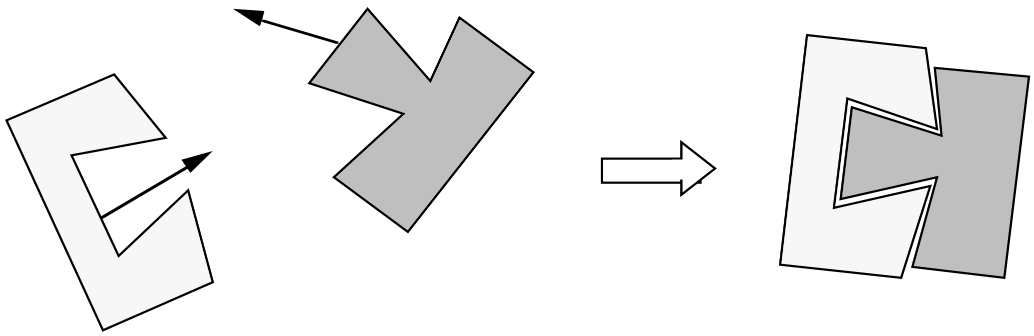
Examples:

- Two magnets attract or repel each other. The north pole of each magnet exerts attractive force on the south pole of another magnet, but repels its north pole.
- Tug-of-war: two groups pull on each end of a rope to try to get the other group across the line.
- A discussion or negotiation between two people. One person says something, the other answers. Argument or proposition is followed by counterargument or counterproposition.

Interaction is a process of variation, but *two* systems are involved. The process can be described as a curve in their shared state space (Cartesian product of the individual state spaces). For example, tug-of-war can be described as a variation of the mid-point of the rope in relation to the line that separates both parties. Interaction is therefore fundamentally equivalent to other forms of variation and should thus satisfy the same evolutionary principles. This means that interaction will in general be subject to selection: certain of the shared states will be fitter and therefore be preferred.

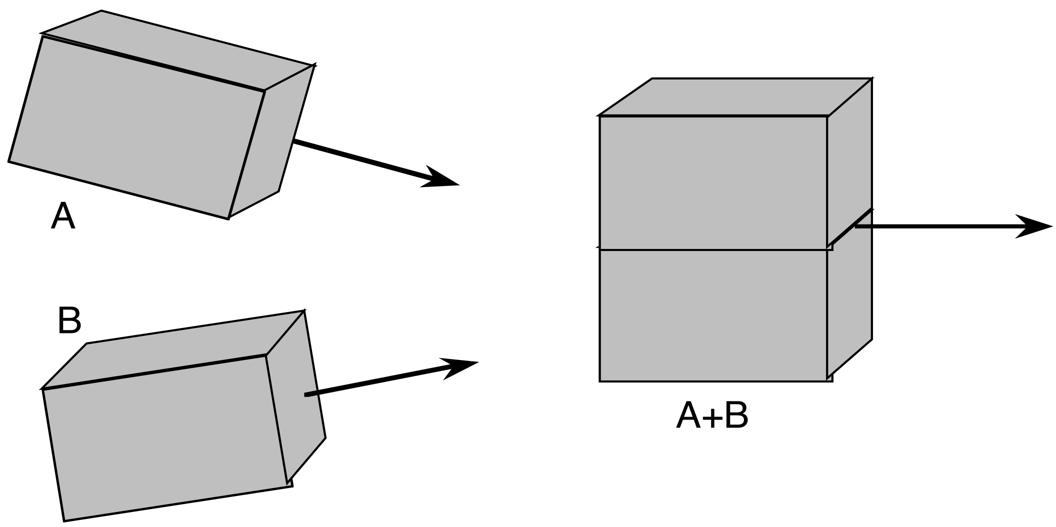
13.3 Bonds

Variation stops when it has reached a shared state (or set of states) that are stable (meaning a valley or attractor in the fitness landscape defined in the shared state space). At that point in time, interaction has been stabilised. The shared state is "fixed". The systems have *adapted* to each other (the one fits in the other).



Examples:

- The two magnets stick together, the north pole of the one to the south pole of the other.
- The discussion ends when both interlocutors agree with each other (or have given up hope to convince each other, they have "agreed to disagree").
- Two symbiotic organisms (e.g. a hermit crab and the sea anemone that grows on its shell) have developed a stable partnership.
- The most traditional example is a molecule, in which the atoms are bound together. Atoms in turn consist of bound elementary particles: protons, neutrons and electrons.



Such a shared, stable state is called a **bond**. The two systems are bound together: one cannot do anything without pulling the other along with it. By definition, the one can no longer vary independently of the other: if one varies, the other has to come along.

Examples:

- If the one magnet is moved, it will pull the other magnet along with it.
- If two people have reached an agreement, and the one does something that concerns this agreement, he or she will have to involve the other.

A bond is a relative **constraint**, that is, a restriction of the freedom of variation in the shared state space, especially the freedom of movement *relative to* each other. A bond thus reduces the freedom of the systems, but increases the predictability of their behaviour.

13.4 System as constraint on subsystem

We can generalise the idea of a bond between two systems to a bond between more systems. For example, a complex molecule binds a large number of atoms, while people in an organisation are bound together by the collective rules that they comply with. In general, these rules mean that the member cannot undertake certain actions without the agreement of the others.

It is the bond or constraint that binds components to a (super)**system**. If the components can vary independently from one another, there is in fact nothing that keeps them together. Then they do not form a system, but an **aggregate**.

Examples:

- Sand is an aggregate of grains of sand. Sandstone is a bond (and therefore a system) of grains of sand.
- People that so happen to walk across the same square form an aggregate. The members of a football team that is playing on that same square form a system. After all, these members depend on each other.

Constraint is what distinguishes a system from its environment: that which follows the rules is part of the system, that which does not, does not belong to the system.

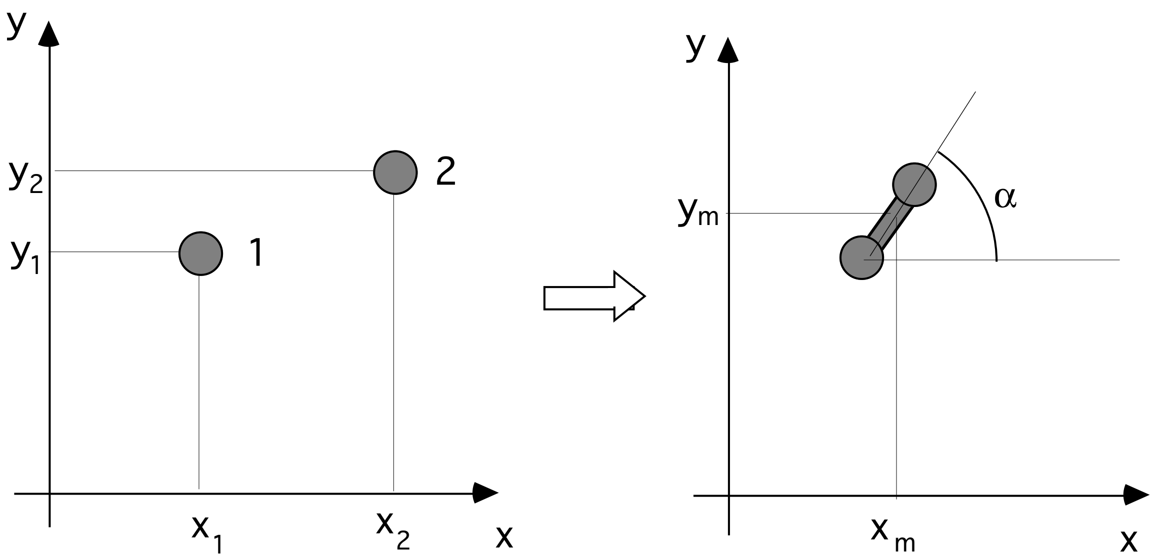
Examples:

- A piece of sandstone clearly stands out from the loose sand in which it lies.
- Individuals that do not follow the rules are expelled from the organisation.

The development of mutual constraint on a collection of loose components is therefore an essential step in the creation of a system. For example, the essence of the transition from loose bricks to a building (system) is bricklaying. A system is a whole with emergent properties. These properties exemplify the relationships between the components, rather than the individual components. It is because of bonds that these relationships become more stable, and therefore more important or exemplary, than the properties of the individual components.

Example:

When we stick two billiard balls together with a rod, the individual balls lose the freedom to take any random position, since the one will have to remain at the same distance of the other. If we want to describe this new system formed by the connected balls (a type of dumbbell), we will in general no longer use the individual coordinates (x_1, y_1) and (x_2, y_2), but the coordinates of the centre of gravity (x_m, y_m), together with the angle α at which the dumbbell is oriented, since these represent the only remaining degrees of freedom. The coordinates of the centre of gravity are not really emergent, since they are determined as the average (deduced from the sum) of the individual coordinates. The angle on the other hand is an emergent property, which is not a simple sum of some individual properties.



Every process of mutual variation and selection of components will therefore sooner or later produce a system. An interesting question is now *how many* components are in general needed to form a system.

13.5 Closure*

It can happen that two interacting systems that reach a bond, have "used up" their interaction capacity. This means that the individual systems can no longer interact with other individual systems.

Examples:

- Na + Cl → NaCl: the NaCl molecule (table salt) cannot absorb any further atoms.
- H_2 + O → H_2O (hydrogen + oxygen = water molecule). The H_2O molecule can still absorb another oxygen molecule, producing hydrogen peroxide: H_2O_2. After this no further reaction is possible.
- Marriage: In general, the bond between one man and one woman excludes other relationships (other than a threesome).

In other cases, new components can still join the bond.

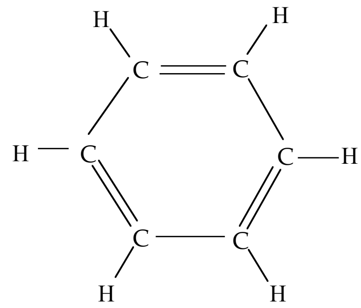
Examples:

- Polymers, i.e. long chains of smaller molecules that keep growing unlimitedly. DNA for example is a polymer.
- Staples that slide into each other and form a forked chain.
- Organisations that continue accepting new members.

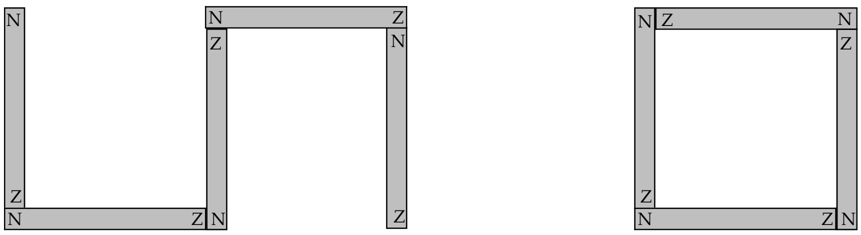
Such "growing" systems continue to have "junctures" to which new components can attach themselves. Each additional component uses a juncture, but adds a juncture in turn. However, it can also happen that the junctures hook into each other, for example, a polymer in which the last molecule hooks into the first one.

In that case, all junctures can be “used up”. All components are completely hooked into each other, without room for further interaction with “outsiders”. In that case, we speak of structural *closure*”: the system has closed itself off and no longer provides access to outsiders.

Examples of closure



- A benzene molecule consists of a closed, hexagonal ring of carbon (C) atoms that are each also bound to a hydrogen atom (H). Other hydrocarbon compounds occur in the form of chains that can grow in length indefinitely by adding extra atoms to the ends.
- In the illustration below, you see bar magnets that are bound together via the attractive forces between their north and south poles. The assembly on the left is “open”: there is still place to add new magnets. The assembly on the right is “closed”: all junctures have been used.



Properties of closure

- Closed systems have the advantage of being more stable: all components are attached at all junctures. With open systems on the other hand, the components at the ends are only partially attached. These can therefore come loose more easily.
- The disadvantage of a closed system is that it can no longer grow.
- A closed system has a clearer distinction between system (inside) and environment (outside), because open systems can still take up components from the environment, or lose components when they come off.

13.6 Selection of stable combinations

Assume that different systems interact with each other. After a while, this interaction produces different bonds. Different combinations of building blocks arise at random. Those that “work” are retained. These building blocks fit together, like pieces of a jigsaw. If the combination is closed, it will no longer grow. Individual components can then no longer interact with other components. A closed system *as a whole* can however still interact with other systems. These interactions are of a different type than the ones between subsystems.

<u>Examples:</u>

- Football players in a team interact by passing the ball to each other; football teams interact by playing matches against each other.

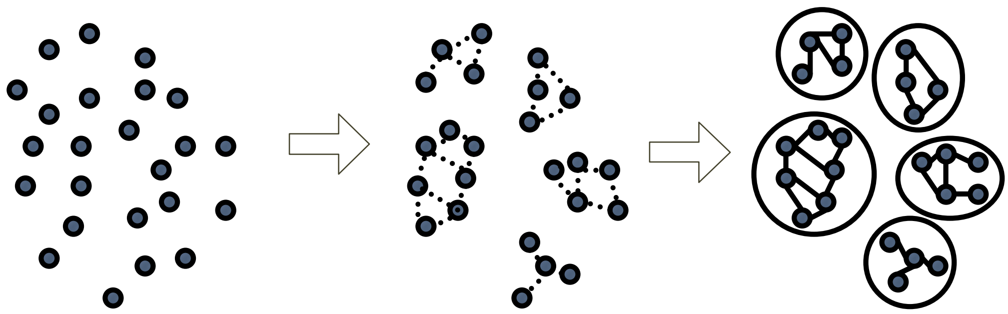
- Protons and neutrons in a nucleus interact through the "strong" nuclear force between particles; atoms interact through the exchange of electrons.

These interactions on the level of the supersystem also lead to bonds between supersystems. This results in a super-supersystem.

These bonds at a higher level are in general weaker than those at the lower level. "Weaker" means less stable, less fit, or easier to break. The reason for this is that evolution prefers the fitter configurations and will try these first before it experiments with the less fit ones. Only when all the "strong" bonds have been realised will variation get a chance to try out the remaining "weak" bonds.

Examples:

- The bonds between elementary particles in an atom are much stronger than the bonds between atoms in a molecule. You need massive particle accelerators that produce a lot of energy to split an atom. To split a molecule, a chemical reaction in general suffices.
- People or animals in a group are easier to drive apart than the cells in their bodies.

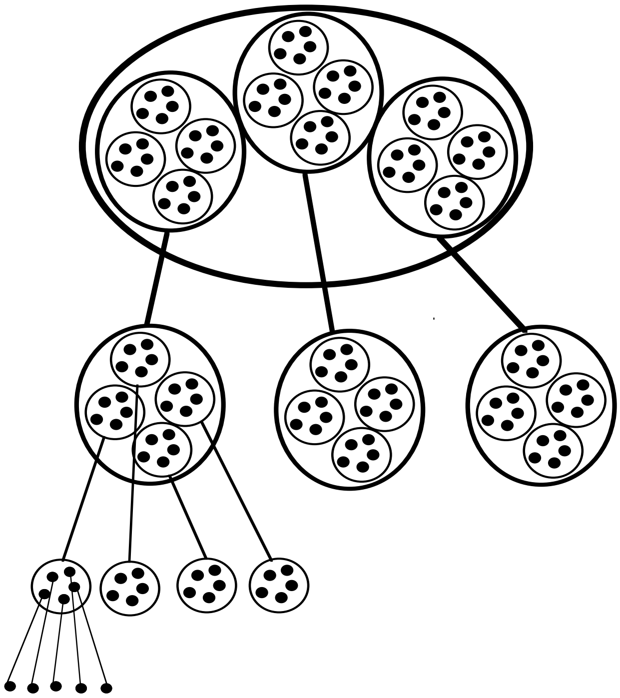


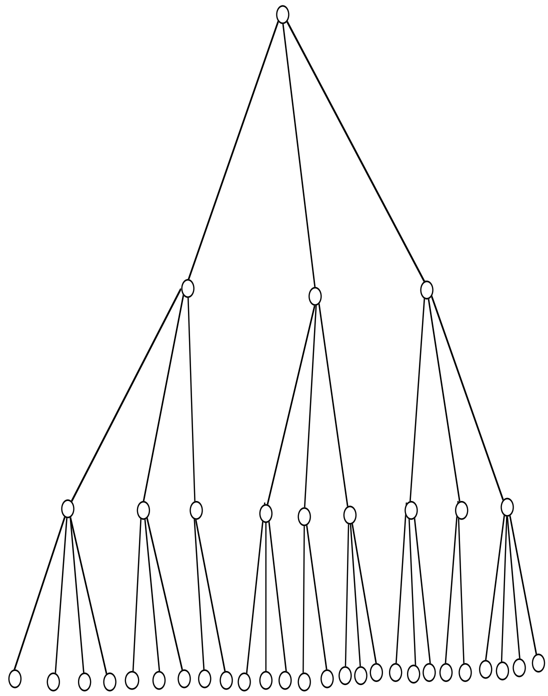
Conclusion: Every system can serve as a building block for a supersystem of a higher order. This supersystem can in turn serve as a building block for a supersystem of an even higher order. In this way, t ever-higher levels of structural complexity arise, for example:

- elementary particles → atoms → molecules → cells → multicellular organisms → societies.
- rocks → planets → solar systems → galaxies → clusters of galaxies.

13.7 Hierarchical architecture

We have seen that evolution constantly generates higher levels of supersystems. The more complex the system, the later its emergence. Each of these systems still consists of all the components out of which it was originally formed. Therefore, every system can be analysed or "decomposed" into its constituent subsystems, which can in turn be reduced to their constituents, etc., to the lowest level, the elementary particles. As we have seen in 4.4, such consecutive layers or levels of "subsystems in systems in supersystems in ..." are called a **hierarchy**.





Such a hierarchy is an example of what in mathematics is called a “tree structure” (imagine the illustration above upside down, with the starting point as the “root” of a tree and the rest as “branches”). This is defined by the fact that every smaller branch (further down) of the tree originates from exactly one larger branch (further up). This means that in a hierarchy every system can only belong to one supersystem.

Note: This is in practice not always the case. A system can be a subsystem of several, overlapping supersystems simultaneously. A person can for example belong at the same time to a family, a football team and a company. This is however rare, because the probability is small that the same state of the system would satisfy different, independent constraints or bonds. Because people are very complex systems, with a very large state space, it is easier for a person to comply with several constraints at the same time (play several “parts” or perform several functions) than it is for an elementary particle.

In practice, the probability of finding a fit combination through blind variation is very small, because there are many more unfit states than there are fit states. An additional constraint on the combination makes it even less likely. Because of that, most complex systems are structured hierarchically. H.A. Simon calls this type of organisation the “architecture of complexity”.

13.8 Nearly decomposable systems

Decomposition into subsystems always remains an *approximation* of reality. Through decomposition we have lost the essential cohesion of the system, i.e. the constraint or the bond that binds the parts into a whole and thus gives the system its identity. At each level, the system is more than the sum (aggregate) of its subsystems. It has emergent properties that were created by the constraint or bond.

Example: A billiard ball is round, and has no orientation or direction. Two billiard balls stuck together form a kind of dumbbell or rod that points into a certain direction. The orientation or “direction” is an emergent property.

In practice, decomposition is still useful. After all, it breaks the weaker bonds that form the supersystem, but keeps the stronger bonds that form the subsystems. The components that remain after decomposition are more stable than the original system.

For those reasons, such systems are called **nearly decomposable** (a concept introduced by H.A. Simon). They are not completely analysable or reducible to their parts, but neither do they form one, indivisible whole. In practice, both reductionists and holists are right: there is a whole that is more than the sum of its parts (holism), but the separate parts still retain numerous fundamental properties and their separation is therefore a useful way to reduce complexity.

13.9 Conclusion: increase of structural complexity

Variation and selection will spontaneously assemble components into higher-order systems, and will subsequently use these as building blocks for systems of the next level. During this process both distinction and connection increase:

- Distinction, because binding or closure of an assemblage of components creates a clear demarcation or distinction between "inside" and "outside"; closure **differentiates** systems.
- Connection, because the components within a supersystem are connected to each other: binding or closure **integrates** the subsystems within a system.

As we noted earlier: differentiation + integration = **complexification**. This complexification in general takes the shape of an ever-growing hierarchy of systems (more levels, more components per level).

Such complexity is however still purely static. Moreover, bonds restrict variation and thus reduce flexibility or mobility. If we consider living beings, intelligence, society and culture, we will find their complexity rather in variability or in adaptivity. To clarify this, we need to introduce another kind of system-building mechanism: the metasystem transition.

Chapter 14. Metasystem transitions

14.1 Adaptivity

All evolving systems "aim" in a certain sense for fitness, meaning that selection implicitly prefers fit systems. It is not necessary for a system to have a built-in goal or plan in order to attain fitness.

Fitness is based on internal stability and external adaptation to the environment. A system that occupies a good "niche" is adapted: it has reached a state that can maintain within the given environment. A subsystem is for example adapted to the system that it is a part of. Once adapted, the system will in general stop evolving: it has reached the bottom of a valley in its fitness landscape.

However, when the environment itself changes, it no longer suffices to be perfectly adapted to the given situation or niche. What was fit at a given moment is in general no longer fit after the environment has changed.

<u>Example:</u> In the snow, it is fit for arctic hares to have a white colour, so that they are less visible for predators. When the snow melts, the white colour becomes much more obvious, and the arctic hare will be eaten sooner.

In a changeable environment, it is useful to be able to adapt *immediately*, rather than having to wait until variation and selection have evolved a new state.

<u>Example:</u> if the warm climate continues long enough, natural selection will give the population of arctic hares a darker fur by eliminating the lighter ones generation after generation. If on the other hand the snow only melts in summer and returns in winter, there is no time to wait for natural selection of the population of arctic hares. In that case, it is more useful for the hares to have this colour change "programmed": white hairs fall out in spring and brown hairs grow in their place, with the opposite happening in autumn.

In this example, the change in the environment is always the same, and therefore predictable. This makes it easier for the genes to evolve a pre-programmed adaptation. In general, however, changes in environment are unpredictable. In that case, adaptation needs to be flexible, rather than pre-programmed.

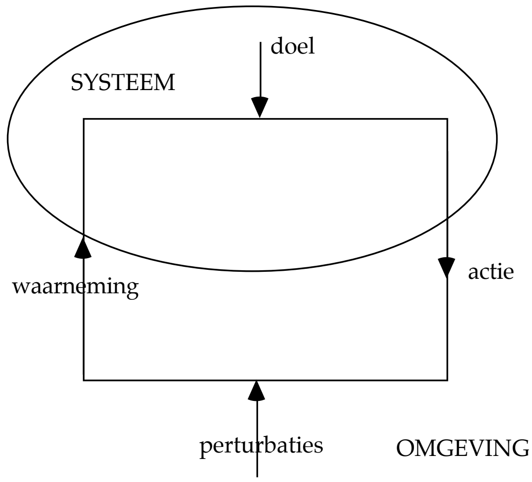
<u>Example:</u> a chameleon or squid can immediately change colour to adapt to the environment, whatever that environment may be.

<u>Definition</u>:

> **Adaptivity** = the ability to change the state of a system in such a way that the changes in the state of the environment are compensated for, and the system remains fit or adapted, despite the changed environment.

In fact, adaptivity is a form of **control** (see 5.3), but now with the goal of fitness. Reminder: control requires a loop with negative feedback, where incoming perturbations (which lead to deviations from the goal state) are observed and compensated for with the correct actions.

Actions can change both the internal state (properties of components, for example skin colour, temperature) and the external state (relationship to the environment, for example proximity to a shelter).



14.1.1.1.1

14.2 Metasystems

Suppose that a system has adaptivity: it varies when the environment varies, but in such a way that its essential properties, its capacity to survive, remain. This means that the variation is no longer random, but controlled. Thus, there has to be some sort of control system present that directs the variations of the system. The combination of the original system (or systems) and such a control system is called a **metasystem**. The original system that is controlled by the metasystem is called an *object system*.

A metasystem is not a static constraint, like a supersystem, but a dynamic constraint, that adapts continuously.

<u>Example:</u> a multicellular organism is more than a supersystem consisting of cells that are bound together. Different cells of different types execute different actions, depending on the situation: for example, muscle cells may or may not contract, nerve cells may or may not pass on signals. These cells are coordinated by a control system, which can be localised (e.g. in the brain), but which can also be distributed (see 3.3) over all the cells (such as the DNA housed in every nucleus). In fact, our body contains a multitude of control systems working at different levels, where the one affects the goals of the other.

Control means that the system is varied in specific ways that depend on the perceived situation. The system undergoes controlled variation rather than random or blind variation. Controlled variation allows the system to maintain its fitness in various circumstances, while avoiding the loss of fitness that is the most probably result of blind variation.

Controlled variation requires in particular that the system have **knowledge** about which variation is appropriate in the given circumstances. As discussed earlier (section 5.8), knowledge can be expressed through condition-action rules of the form: IF a certain situation is observed, THEN execute a certain, adapted action. In the short notation: condition → action.

14.3 Hierarchies of metasystems

Just like a supersystem is in general comprised in a higher-order supersystem, so too a metasystem can be subject to a higher-order metasystem, and so forth. This leads to a generalised control hierarchy (section 5.6), which is different from the structural hierarchy that we have discussed above (13.7). The simple formula for this is as follows:

metasystem = control of object system

The next level in the hierarchy is then described as:

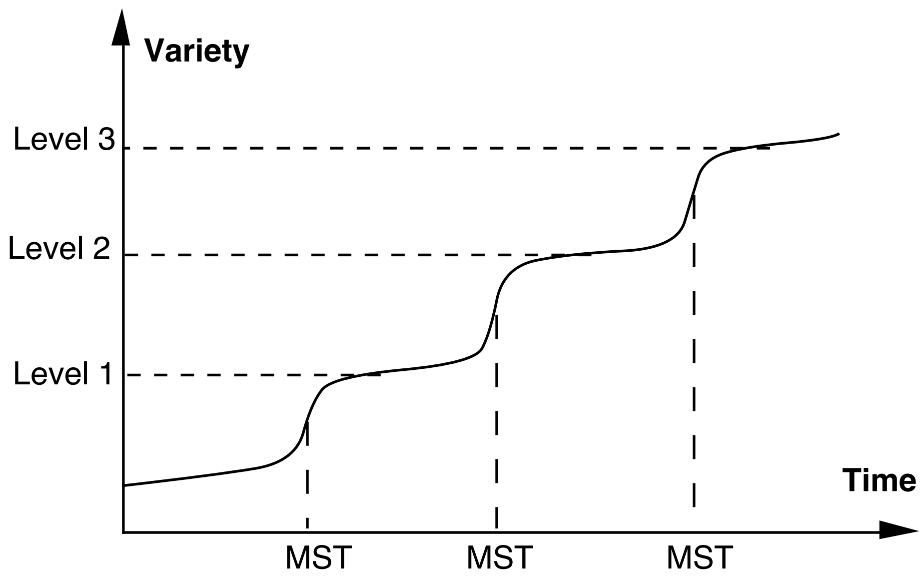
meta-metasystem = control of metasystem, etc.

Example: thermostat = control of the heating; presence sensor = control of the thermostat.

This can also be expressed using condition-action rules. The metasystem of the first level varies the state of the object system according to certain rules. The rules for the metasystem of the second level then determine the variation of the rules of the first level.

Example: The thermostat varies the amount of heat that is produced according to the rules: temperature lower than goal temperature → turn heating on; temperature high enough → turn heating off. The presence sensor varies the setting of the thermostat, and therefore the way in which the thermostat regulates the heat influx, according to the rules: someone in the room → set goal temperature to 21º; room empty → set goal temperature to 16º.

Just as evolution spontaneously leads to higher levels of supersystems, so too does it lead to higher levels of metasystems. The emergence of such a higher level is called a **metasystem transition** (MST, a concept introduced by V. Turchin). The reason why variation and selection produce such transitions is simple: a metasystem has a larger adaptivity (and therefore a greater variety of possible states and actions) than an object system, and will therefore be fit in more diverse circumstances, but a meta-metasystem has an *even* larger adaptivity. Thus each MST produces a big jump in the system's variety of actions and adaptivity.



While a metasystem will always react in the same way to a given condition, a meta-metasystem will be able to adapt its reaction by taking into account additional conditions, which are typically of a higher, more abstract level. A meta-meta-metasystem will in turn have a larger adaptivity than a meta-metasystem. Each mechanism of adaptation can itself become subject to an adaptation mechanism on a higher level. The larger the adaptivity, the larger the (internal or absolute) fitness, and thus the greater the chance that such a system will sooner or later evolve.

Limitations of metasystem hierarchies

In spite of these advantages, a hierarchy of metasystems is far more complex in its functioning than a system with one or two levels. To determine which action is to be executed, the decision has to pass through different control levels, which each have to check their own conditions to potentially change the setting of the underlying control level. For this, each level needs to receive information about the state from lower levels, process and interpret this information, and send the decision back down to the lower levels. A large number of control levels thus slows down the decision about which action will eventually need to be executed, and increases the chance of errors or miscommunication between the levels. Adding metalevels therefore does not only have advantages, but also disadvantages.

Example: Complex bureaucracies with many administrative levels, such as in ministries or in large companies, are in general slow and rigid. That is why it is fashionable at present to flatten hierarchies and decrease the number of levels.

The solution is to make each control level as intelligent and autonomous as possible, so that it can make as many decisions as possible on its own, without needing to involve the higher levels. No more hierarchy should be instated than absolutely necessary. The cyberneticist Aulin has formulated this principle as the *law of requisite hierarchy*: the more control the lower levels have, the less hierarchical levels are needed to gain control over the global situation (and vice versa).

Example: the reduction of hierarchy in a company or administration is only possible if the employees at the different levels are sufficiently intelligent and informed to work largely on their own, without requiring constant supervision or orders coming from above. This is one of the reasons why "flat" organisation structures are much more common in the current age, with its efficient information technology and well-educated employees, than a century ago.

In evolution, the problem of too many metalevels does not appear: the creation of a new metalevel through variation and selection is very difficult and will only occur if there is no other way left to increase fitness. In a bureaucracy, on the other hand, it is easy and tempting to continuously add new levels, so that managers or civil servants can always be promoted to a higher level. In evolution, on the other hand, the existing level will first be extended and made as efficient as possible before a new level will be tested out.

The autonomy of the lower-order systems is maintained as much as possible. This means that most actions will by default be undertaken by the lower-order systems. These are after all very well adapted for this, after hundreds of millions of years of variation and selection. The higher-order metasystem will only intervene if the situation becomes too exceptional or complex to be solved by the lower-order system alone. In this way, the delays and confusion that are likely to appear if you would go through all levels of the hierarchy are avoided in the overwhelming majority of cases.

Example: Our respiration and our reflexes (such as pulling away your hand from a hot surface) are largely controlled by a lower-order, subconscious system, which works fast, efficiently and automatically. In special circumstances, our conscious thought (a higher-order system) can intervene and make the lower-order system deviate from its normal settings. For example, if necessary we can hold our breath (for example if we know there is a poisonous gas present) or force ourselves to walk over hot coals. Our consciousness would however become overloaded if it would have to reflect about every heartbeat or breath.

14.4 The metasystem hierarchy of Turchin

The Russian/American cyberneticist Valentin Turchin has formulated a sequence of the most important metasystem transitions in the evolution of life, from primitive animals to human culture. We will now discuss it briefly.

- Simple reflex = control of movement

This first level is characterised by *reflex movements*, where a specific perception by a sense organ immediately activates a certain muscle via a connecting nerve. Although we still have such reflexes, this level of control is typical for primitive animals, such as sea anemones or worms, that have nerves, but do not have brains. These animals always react in the same way to the same simple stimuli. For example, touch → sea anemone contracts.

- Complex reflex = control of simple reflex

When different nerves, coming from different sensors (sensory cells) come together, the organism can weigh the importance of the different stimuli or sensations against each other, and produce an integrated reaction that considers the different aspects of the perception.

<u>Example:</u> If the sensation "touch" is combined with the visual perception "large" or "small", the animal can react in a more adapted manner: "touch" by something "large" → danger, pull away; "touch" by something "small" → prey, eat.

This level of control assumes a crossroad of nerves, where the different stimuli come together and are compared, before they are passed on to the muscles that execute the actions. This junction forms a rudimentary brain.

- Learning or associating = control of complex reflex

Complex reflexes are still rigid, and the same combination of stimuli will always lead to the same reaction. However, when the environment is complex and changeable, it is useful to adapt one's reactions to new phenomena, that is, to **learn** new condition-action rules or condition-condition rules. Knowledge no longer needs to be inherited, but can be learned individually. Learning is therefore a controlled variation of rules.

Learning works through *association*: if an observed condition A is frequently followed by a condition B, the organism learns to make the association $A \rightarrow B$. This means that if A is observed, it creates a more or less strong *expectation* of B. The more often the two conditions are observed together, the stronger the association or expectation, and the stronger the chance that the rule $A \rightarrow B$ will be effectively applied to predict what will happen.

<u>Example:</u> Pavlov's experiment: if a dog always gets food just after a bell has rung, the dog will begin to salivate when it hears the bell, because it expects food.

Learning also works through reward / punishment of actions (*reinforcement*): if a condition A is followed by an action B, and the result is positive (the organism comes closer to its goal state), then the association $A \rightarrow B$ is reinforced or rewarded. If on the other hand the result is negative, the association is weakened or punished.

<u>Example:</u> A rat that perceives a lever and pushes it, is always rewarded with food. The rat will soon learn to keep pushing the lever. However, if that same rat gets an electrical shock when it pushes the lever, it will learn to stay away from it.

- Thinking = control of associating

Through learning by association, you can only make connections between phenomena that are *observed together*. Phenomena that have never been observed together are not associated with each other. Thinking means that you can conceive combinations of phenomena that you not necessarily have experienced in reality. You can thus learn rules without having experienced them. This is typical for human intelligence, and is absent in animals.

Example: We can imagine an elephant wearing a top hat, without ever having seen one, or we can consider how we would solve a problem (such as engine failure in a motorboat) that we have never experienced.

Thought uses concepts represented by symbols (such as words) that can be combined according to certain combination rules (such as grammar and logic). Because the concepts are abstract or symbolic, they do not depend on concrete perceptions. This allows for free combination, independently of what is happening in the environment. The combination rules ensure that this process of variation is still controlled to a certain extent.

- Culture = control of thinking

We have in general not personally invented the words, concepts and reasoning rules that we use, but copied them from others, from the ideas that were proposed to us in society and during our education. This dependency limits us in the freedom of our thought, in our creativity. New concepts and rules in general arise through a process socio-cultural evolution, over which an individual has no control. Until now, this cultural evolution is still mostly a process of blind variation and natural selection of the most useful concepts. It is not directed, and therefore not very efficient.

As society and culture become more complex, tools, systems and methods arise for the systematic discovery of new concepts. Examples are the scientific method, philosophical analysis, artistic exploration, computer programs for the discovery of patterns in data, communication technology for the exchange and discussion of ideas... This development appears to go faster and faster. It seems as if we are on the verge of a new metasystem transition to a higher level of organisation. A plausible conception for this higher system level is the *global brain*, i.e. the intelligent system that emerges from the integration of all people and computers on this planet. The Internet here plays the part of the nervous system of this supersystem, storing and propagating information, but also supporting the development of new concepts and theories.

14.5 The origin of life*

Probably the most fundamental metasystem transition is the origin of life itself. This is absent in Turchin's sequence. Living beings are systems with in-built adaptivity. This means that they are purposive, with survival or fitness as their fundamental goal. To understand how life could emerge, we have to ask ourselves how variation and selection could have produced goal-directed or control systems. The following considerations are mostly speculative, but offer a plausible picture of this most mysterious transition of all.

In general, it is assumed that for the most primitive possible organism (a type of rudimentary bacterium) three components were needed:

1. A membrane or cell wall, which protects the system from the environment, but lets in food.
2. A "*memory molecule*" or "replicator" (such as DNA) that "remembers" the operation of the organism and that can be copied (replicated), and in this way passed on to descendants, possibly with variations.
3. An *autocatalytic cycle* of chemical reactions that uses incoming food molecules to produce, repair or reproduce internal components.

Catalysis is a term from chemistry. It refers to the process where a certain type of molecule (the catalyst) enables, facilitates or speeds up a chemical reaction between other molecules. Autocatalysis then means self-enhancement or self-facilitation. This means that the molecules in the cycle stimulate their own production. Autocatalysis is a form of positive feedback.

<u>Examples:</u>

- $A + X \rightarrow 2A$

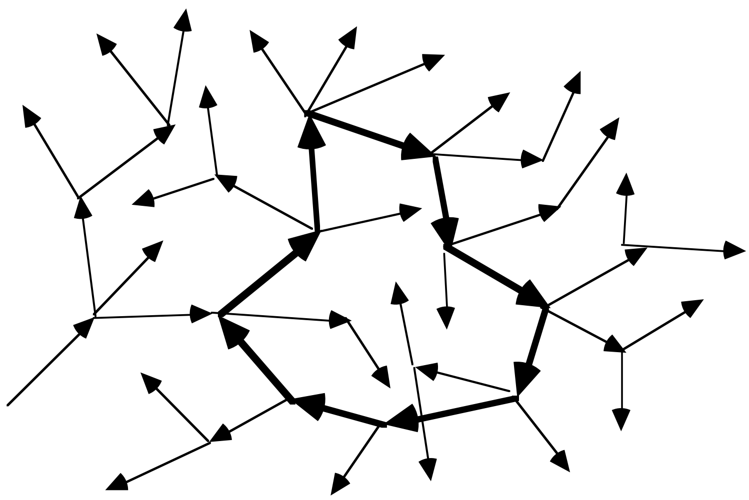
  A doubles itself in this reaction. This is a cycle because you start and end with A.

- $A + X \rightarrow B$, $B + Y \rightarrow C$, $C + Z \rightarrow 2A$,

  $2A + 2X \rightarrow 2B$, ... $\rightarrow 2C \rightarrow$ ...

  This more complex cycle consists of three intermediary steps (involving respectively A, B, C), after which we return to the starting point (A), except that we now have a double amount. X, Y, Z play the part of the "food molecules" that are being used to produce more As, Bs, and Cs.

Scientists do not yet agree which of these three components is the most fundamental, and how, or in which order these components have originated. If we consider a living being as a control system, we see that it does not matter in which order the components originated, as long as they have been integrated at some point. It is for example possible that empty membranes, free "replicators" and autocatalytic cycles have developed independently, in parallel. Membranes that happened to include an autocatalytic system would probably have functioned better, because the molecules produced by the cycle helped the membrane to grow, while the membranes in turn would have protected the cycle from disturbances. "Replicators" that happened to be part of an autocatalytic cycle would have done better for the same reason, while they would have stored the operation of the cycle in their "memory", so that this cycle would have been reproduced more easily.



- How can each of these components have arisen on its own?

- *Membrane*: certain simple fatty molecules (bilipids) self-organise spontaneously in the shape of two-dimensional layers, which close in on themselves and in this way form "cells". When the number of bilipids increases, these cells can spontaneously split in two.

- *Autocatalytic cycle*: If the network of reactions is complex enough, and each molecule produces one or more new molecules, which in turn produce new molecules, etc., a loop will sooner or later always develop, meaning that one of the input molecules appears again as part of the output. The necessity of the appearance of such a cycle has been argued by Stuart Kauffman, using computer simulations and mathematical reasoning, but is in fact a rather obvious mechanism. The thick arrows in the illustration below represent a cycle within a larger, random network.

  Such a cycle is a form of positive feedback, but also of closure, as described in 13.5, albeit no longer static, but dynamic. Moreover, as per the definition of catalysis, more will now be produced than when the chain was begun. All molecules in the chain are therefore increasing in number, as long as there are enough "food" or "fuel" molecules present to allow the reaction to continue.

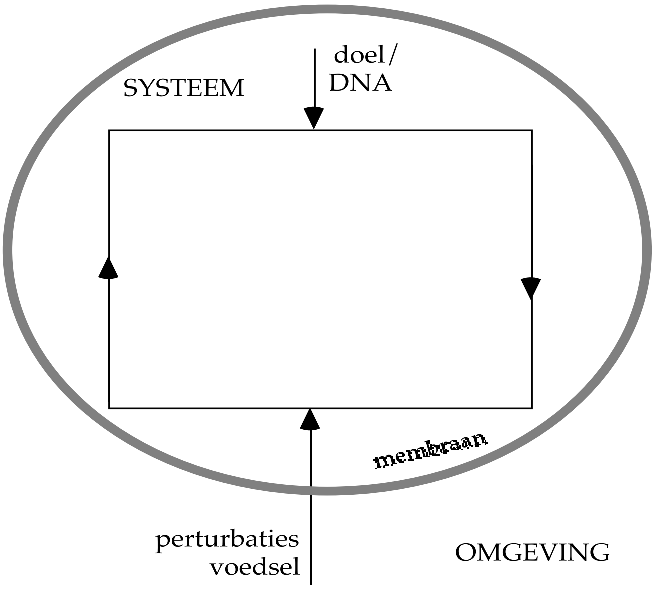
- *Memory molecule*: The simplest known type of such a molecule is RNA, a single form of the "double helix" of DNA. Relatively simple forms of RNA have been found that are capable of reproducing themselves. At the moment, there is not yet a concrete scenario for the development of the first RNA, although the components of RNA are relatively easy to produce. RNA may have developed as a by-product of an autocatalytic cycle, as a more complex form of an originally simpler molecule that no longer exists, or (unlikely) directly through combination and selection of its components.

14.6 Origin of control*

We are now looking for a scenario that allows a goal-directed, adaptive system to arise spontaneously from non-goal-directed components. This means that we will try to explain the origin of life using evolutionary and cybernetic principles, rather than the chemical properties of molecules and reactions.

- What is the minimum we need for a control system?

1. **Negative feedback**: ensures that perturbations are suppressed and that the system returns to equilibrium.
2. A **goal** that represents the preferred state: it ensures that the state to which the system returns is indeed the fittest one, the one that will lead to long-term survival.
3. *Amplification*: the effect of actions is greater than the effect of perturbations. This ensures that the suppression is strong enough so that deviations caused by disturbances never become large enough to endanger survival.



Such a rudimentary control system for a living cell is illustrated above. Note that this diagram is equivalent to the general diagram for a control system (5.5), except that the perceptions and actions now take place *within* the system boundary formed by the membrane. The reason for this is that most primitive cells do not yet have external sensors or organs with which they can interact directly with their environment. They have to wait until the environment itself enters the system, and suppress (or possibly reinforce) the effects of this intrusion.

All three "life components" have their own function or role to play in the construction of such a control system:

- *memory molecule*: storage of the **knowledge** and **goal** which direct the autocatalytic cycle. This means that they ensure that the right molecules are produced in the given circumstances.
- *membrane*: **buffering**, i.e. the passive absorption of disturbances from the environment. This weakens the effect of disturbances on the goal, and thus contributes to amplification. Moreover, it filters incoming molecules, so that in the first place only "good" (food) molecules can enter, and "bad" (toxins) cannot.
- *autocatalytic cycle*: production of sufficiently strong **actions** to compensate for disturbances and rebuild damaged components. This also contributes to amplification, but requires energy in the form of food.

The autocatalytic cycle is however a *positive* feedback loop: you end up with more than you started with. Without control, such a cycle would lead to a run-away growth. Such growth requires the availability of food molecules, though: when these are exhausted, the cycle stops. The food molecules in turn depend on what is available in the environment. This makes the operation of the cycle very unreliable: depending on availability, it runs either too fast, or it is standing still. What we need is a combination of these mechanisms with negative feedback and thus control.

- Possible scenario for the construction of a control loop from the components

Imagine that one of the "food molecules" necessary to allow an autocatalytic cycle to function is not coming from the environment, but from the cell itself. The molecule can then limit the speed of the cycle independently from the environment. After all, the cycle cannot produce anything without this molecule's cooperation. The molecule then plays the role of a "goal molecule" that controls the cycle. As long as there is more food than goal molecules available, the speed of the autocatalytic reaction will depend only on the goal molecule. The least available molecule after all restricts the production of all others (it is the bottleneck of the process). If this molecule increases in number, the speed increases; if it decreases in number, the speed decreases. In this way, a "goal molecule" can completely control the reaction.

Now imagine that there are several types or variants of the goal molecules, that each send the cycle into another direction, meaning that different molecules will be produced. Then the goal molecule does not only have control over the *number* of molecules that are produced, but also over the *kinds* of molecules. This means that the produced molecules can be adapted according to the specific circumstances.

Example: different types of food require the production of different types of "digestive enzymes", and this requires that the goal molecules send the cycle in the right direction to produce these enzymes.

Such adaptivity can be achieved by having goal molecules produced by a separate autocatalytic cycle that is influenced by what is happening in the environment. Such a cycle that controls one or more other cycles is called a *hypercycle* (a concept introduced by molecular biologist Manfred Eigen). The development of a hypercycle is an example of a metasystem transition. Once such a hypercycle has emerged, we get an organisation similar to the one in a living cell.

Example: In a cell, certain pieces of DNA are "activated" by other molecules (under influence of the environment). The activated DNA produces specific RNA molecules. The RNA molecules catalyse or direct specific reactions, adapted to the circumstances.

Chapter 15. Conclusion: the evolutionary-systemic worldview

15.1 Ontology

We will now summarise the philosophical implications of the previous chapters in order to survey the overarching **evolutionary-systemic worldview** (ESW). Every philosophical system is based on an **ontology**. This answers the questions "what is?", "which are the fundamental categories of existence?"

In the ESW, the fundamental components are relational: distinctions and connections. Distinctions and connections are the building blocks for more complex structures or forms of organisation: systems. Systems arise spontaneously (self-organisation) through variation and selection of combinations between simpler component systems. Initially, we only see simple systems, out of which more complex systems evolve gradually. Systems are selected for their fitness: the ability to survive and be produced or reproduced. Fitness can be reached through: 1) stability (internal); 2) adaptedness (external); 3) adaptivity.

Once a certain degree of complexity is reached, systems can develop goal-directedness. They are no longer passively subjected to their environment, but autonomously aim for their own goals, by intervening in the environment if necessary. Goal-directed systems have adaptivity: if the environment changes, they can adapt immediately, without having to wait for evolution to generate a new design. In order to be able to do this, they need knowledge: instead of having to apply blind variation (trial and error), they "know" what to do in given circumstances.

Systems form different hierarchies of complexity, with the system at each higher level more complex than the one at the lower level:

- **Supersystems** (static): atoms, molecules, rocs, planets, etc. Supersystems consist of different components or subsystems (distinctions) kept together by connections (constraint).
- **Metasystems** (dynamic, goal-directed): living beings, intelligence, society, ... Metasystems consist of a control mechanism that coordinates one or more object systems while directing them at a certain goal. For metasystems that have developed through evolution, the ultimate goal (highest metalevel) is always fitness. Goals on lower levels are subordinate or instrumental and can vary widely based on survival strategy, system, environment and circumstances.

15.2 Epistemology

Epistemology is the branch of philosophy that is concerned with knowledge and that wonders how "true" knowledge can be distinguished from "false" knowledge.

According to the ESW, knowledge is a product of evolution: it does not come from an abstract realm of "Ideas" as Plato thought (*idealism*), but neither is it the result of a mere passive observation of the environment (*empiricism*) that produces an objective representation or reflection of reality (*naive realism*).

Knowledge arises through blind variation and selection of potential knowledge structures. Knowledge must be constructed by the system itself, by trying out various possible rules or combinations of rules, and seeing what works. The environment here plays the role of selector (*selectionism*), which eliminates or penalises bad rules, but does not tell the system what the good rules would be. The system receives no *instructions* from the environment, the way pupils are instructed by their teacher. It is responsible for the construction of its own knowledge (*constructivism*).

<u>Example:</u>

Imagine a small ocean-living organism, which, in order to survive, needs to stay within the right temperature zone. It can achieve this by moving up to warmer or down to colder water layers. The organism distinguishes three conditions or states of the World:

$$W = \{\textit{too cold}, \textit{exactly right}, \textit{too warm}\}$$

It can perform three Actions:

$$A = \{\textit{go up, go down, stand still}\}.$$

The knowledge structure of the organism consists of a function or set of rules that maps each of the elements of the set W onto an element of the set A:

$$f: A \rightarrow W.$$

There are $3^3 = 27$ possible combinations of such rules, but the only real fit one consists of the rules *too warm* → *go down*, *too cold* → *go up*, and *just right* → *stand still*. All other 26 combinations (e.g. *too warm* → *go up*, *too cold* → *stand still* and *just right* → *go down*) will sooner or later get the organism in trouble. Therefore, they will be eliminated by natural selection. There is therefore a probability of 1 in 27 that the organism will find the correct combination through blind variation of possible combinations.

More realistic organisms on the other hand have thousands of possible perceptions and thousands of possible actions. This makes the number of possible combinations of rules astronomically large. In this case, selection by the environment on blind variation will no longer be enough to find the correct model. An organism will then have to appeal to internal selection criteria or rules, that make a "pre-selection" of plausible rules (see 12.4). These internal rules play the part of selectors, which take the place of natural selection by the environment. The evolutionary epistemologist Donald T. Campbell calls these *vicarious selectors*. These internal selectors themselves, however, have developed through blind variation and selection. The development of such a selector is an example of a metasystem transition. Thus, vicarious selectors will be organised in a control hierarchy, with the higher levels varying and selecting the lower ones.

Knowledge is not an objective reflection of the environment, but a tool for control. This means that knowledge helps a system to attain its own, subjective goals. Knowledge selects for the right actions (or rules for actions) *before* selection by the environment gets the chance to eliminate a system. Different systems with different goals or different strategies to adapt will therefore have different forms of knowledge, even if they live in the same environment.

15.3 Ethics

Moral philosophy or ethics is the branch of philosophy that is concerned with the behavioural rules that we need to follow to constructively live in a group, without harming others. As we have already alluded to several times (2.3, 6.7), the theory of evolution offers a profound explanation for why such rules are necessary, and how they have developed.

- The evolution of altruism

There is always competition for scarce resources. Not every individual can be successful, so only the fittest are selected, independently of the others. Helping others in general does not contribute to one's own survival. On the contrary, it costs energy or effort, while it will in the first place benefit a competitor, who after all needs the same resources. This is why natural selection will produce selfish or egoistic individuals, who only stand up for themselves. The opposite of selfishness is *altruism*, meaning behaviour that benefits others rather than oneself. An example is jumping into the water to save someone who is drowning, at a danger to one's own life. The problem then is: how can altruism, which undeniably exists among both humans and animals, have evolved?

One obvious solution is *group selection*: a group in which the members help each other will in general survive better. For example, wolves that cooperate can kill much bigger prey (e.g. elk), than if they would hunt alone. For this reason, altruism within a group benefits its members. This is why there will be a selection of groups consisting of altruists, at the expense of groups with only selfish individuals.

However, there is a fundamental problem with this mechanism: selfish individuals within an altruist group receive more benefit from altruism than the altruists do. They let others do the dirty or risky work; they do not contribute to the costs, but they do reap the rewards. For example, "free riders", who do not pay for the subway profit from others' contributions to maintain the public transport system. In this way, the selfish ones within an altruist group will be fittest. In this way, the tendency towards altruistic behaviour will be selected against in the long run.

So how can altruism, cooperation or morality evolve? Sociobiologists have suggested two fundamental solutions for this problem:

- *Kin selection*: selection based on kinship or blood ties. Helping family members, especially one's own offspring, benefits one's own genes. A gene for "nepotism" (favouring family) will therefore be selected for. This explains among other things the care of parents for children, and the solidarity between siblings, but also the very complex cooperation in insect colonies. All bees, ants, etc. in the colony are after all descendants of a single "queen", so that everyone is closely related to everyone.
- *Reciprocal altruism*: the tit-for-tat strategy (see 6.7): I help you on the condition that you later help me; if you do not return the favour, I will stop helping you. This explains for example the solidarity between vampire bats (see 2.3). To allow this to succeed you need trust in others, and this requires familiarity with one another. This is difficult in large groups in which you know only a few people, so that you cannot build up a relation of trust. For example, a shopkeeper can easily deceive tourists, because these will most likely not return to the same shop.

These mechanisms are however not sufficient as explanation for the very sophisticated and extended solidarity and cooperation in human society. We will therefore not only need to study the evolution of morality as a biological phenomenon, but also as a cultural phenomenon.

- The cultural evolution of morality

Not only genes, but also memes (ideas, norms, traditions) are subject to evolution. This in general leads to a *conformist* distribution within a group. If someone has a good idea, it is advantageous to copy that (cf. 6.8). Mutual altruism is a good idea and will thus be imitated more and more. The more people do something, the more that behaviour is imitated. Eventually, *everyone* in the group will exhibit the behaviour (conformism), and no one will be able to evade that conformist pressure. Consequently, selfish individuals are sidelined, because they can no longer profit from the efforts of others.

This then causes cultural group selection. If everyone is (mutually) altruistic, one-to-one reciprocity is no longer needed to keep out cheaters or free riders. The idea that remains is altruism towards everyone in the group. However, different groups with different forms of mutual help and solidarity now enter into competition with each other. Those with the best cooperative system will be most successful. As a result, their system will eventually be adopted by the less successful groups. In the end, everyone will have a similar morality.

Note that not all rules of such an evolved morality will effectively be advantageous. Natural selection only eliminates "bad" ideas, such as human sacrifice, that decrease the fitness of the group. Neutral or irrelevant ideas tend to persist. Rules that "happened" to work well, for example, not eating pork (Islam) or beef (Hinduism), will continue to be obeyed even when they no longer have a function, because of conformist pressure making it impossible to deviate from them. This evolution in general leads to complicated systems of gods, precepts and traditions in different cultures. Moreover, the ideas that used to work well in the past do not necessarily still apply to the present. The environment has after all changed. The mechanism of conformism is conservative: it is very difficult to initiate a new idea that goes against established ideas. This is why morality cannot be based on tradition alone—although tradition does offer a useful starting point, because the traditional rules have undergone an extended selection that has eliminated rules that did not work.

- The problem of ethics

The problem remains how we can found a system of ethics on a more scientific basis. How can we optimally coordinate the fitness of the components (individuals) and the whole (group, society, ecosystem)? This problem is not trivial, because what is best for a subsystem is not necessarily best for the system as a whole—which after all has emergent properties. For me personally, it is for example better not to install a filter on the exhaust of my car, because that brings additional costs without being particularly advantageous for me. For the world as a whole it is however better that all exhaust gases would be filtered so that pollution is minimized.

You need a thorough evolutionary-systemic analysis to reconcile the values of the individual or subgroups as best as possible with the values of the society as a whole. Such an analysis will hopefully allow us to develop a morality that does not only benefit as many people as possible, but that is also easily realised in practice, without requiring too many heavy-handed control mechanisms, such as police, courts of justice, and prison sentences to punish offenders.

15.4 Answers to the fundamental questions

Let us conclude this book by summarising the ESW in the form of answers to the main existential questions that all worldviews should address (1.1). The answers here are necessarily short and simplified, but together with the previous concepts and principles, they hopefully offer an acceptable view of our place in the cosmos.

- Why is the world the way it is?

The present state of the universe is partly the result of accident (because variation is intrinsically unpredictable), partly the result of rationally understandable regularities (because the concept of "fitness" and of its derivates, such as closure, allows us to a certain extent to predict which variations will be selected).

- Where does the world come from?

Evolutionary principles and concrete observations allow us to reconstruct how all fundamental systems have evolved from one by one: Big Bang, elementary particles, atoms, molecules, stars, planets, cells, multicellular organisms, animals, humans, society, culture...

- Where do we come from?

Humans evolved from animals that had the capacity to learn, by undergoing yet another metasystem transition to the level of thinking. All details of this transition are not yet known, but it probably has to do with the origin of a grammar-based language among chimpanzee-like proto-humans. Chimpanzees and their relatives (bonobos, gorillas) exhibit a great similarity with humans in fundamental biological, social and even mental aspects, including the use of tools, primitive language and culture. One or several accidental factors (such as a climate change that transformed the rain forest to a savannah) probably caused initially small lifestyle changes (such as walking upright, thus keeping hands free to use tools, and switching from a mainly plant-based diet to a more meat-based diet thus providing the extra energy needed to support a larger brain) that suddenly sped up the evolution of tool use, language and culture. The more complex language and culture led to selection for higher intelligence (and thus a larger brain) in individuals. This increasing intelligence in turn stimulated the further complexification of language and culture. In this way, cultural evolution and the brain evolution boosted each other in a positive feedback loop.

- Who are we?

At the moment, humans exhibit the highest metasystem level. This gives them unprecedented power and insight relative to other organisms, and unique properties such as imagination, creativity, abstract thought, self-awareness, etc. However, they still have plenty of the limitations that they have inherited from their ancestors, such as aggression, faulty reasoning, jealousy, chronic stress reactions, etc. These reactions may have been adapted to the environment of our ancestors, but in our current environment they are rather counterproductive. A better understanding of evolution will allow us to become more conscious of these limitations and in that way to deal with them better.

- Where are we going?

The ever-faster evolution of science, technology and culture appears to herald a new metasystem transition. This will lead to a system with as yet unpredictable capacities for adaptation, creativity, thought, consciousness and action. Probably the best metaphor for this is the “global brain”, the thinking system that arises through the integration of all individuals on this planet via an intelligent computer network.

- Who am I?

Each individual is a unique combination of genes (except for identical twins) and experiences. Although the variety of the number of possible human individuals is infinite, it is not unlimited: there certainly are forms that are “unviable” or “inhuman”. This means that the variation between people obeys certain constraints. It is thus possible to define an infinite “state space” of possible personality types, in which everyone can recognise their own, unique, properties.

Psychologists have used a large-scale statistical analysis to determine a space with five fundamental personality dimensions, the “big five”. They are:

- introversion—extraversion,
- emotionality—stability,
- openness—conservatism,
- agreeableness—disagreeableness,
- conscientiousness—impulsiveness.

Each has an evolutionary meaning which in itself is neither good nor bad, but of which the fitness depends on the environment (for example, “open” personalities will aim to gain more experiences, but will therefore also run more risk in dangerous environments; agreeable characters will fit better in an cooperative group, but will be taken advantage of in a selfish group).

- Where am I going?

Every individual will sooner or later find their own niche, suitable to their unique personality. This means that people during their development will make different choices about their education, hometown, partner, friends, profession, ... until they, through trial and error, reach a situation in which they feel good. Psychological research shows that the eventual niche depends to an important on genetic background: identical twins that were separated at birth and thus have been raised in different ways, still tend to resemble each other in their life choices: independently of each other, they often end up in a similar career (such as secretary or fireman). The better you get to know your own personality, the easier it will be to find a suitable niche and develop yourself further. Thus, an introvert-stable-conservative-conscientious character will probably be happy as an accountant, while an extravert-emotional-open-impulsive character will rather have a career as a rock singer or actor, and an introvert-emotional-open-conscientious character as painter or poet.

- Is there an ultimate goal?

No, evolution does not aim for a final goal, as postulated by the evolutionary theologian Teilhard de Chardin, who called it the Omega Point. Evolution will never be finished. Although evolution produces goal-directed systems, it is itself not goal-directed. Evolution remains intrinsically unpredictable. Still, this evolution is not random, but has a preferred direction: the increase of fitness.

- Does a God or higher power exist?

From the evolutionary viewpoint, an explanation going back to some Creator who rules the universe is not only superfluous, but misleading rather than enlightening. If you wish, you can however see the process of evolution itself, or the universe that was produced by it, as “divine”, in the spirit of pantheism. In that sense, you can have spiritual or religious feelings (connectedness to the whole, awe in the face of the unimaginable complexity of the universe, attraction to the mystery of everything that has not yet been explained, trust in the power of evolution to always get out of problematic situations), without believing in a personal God. The philosopher Leo Apostel called such an attitude “atheistic religiosity” or “atheistic spirituality”.

- What are good and evil?

There is no absolute good or absolute evil: what is good in certain circumstances (for example having sex with someone you love), can be bad in others (for example having unprotected sex with someone you do not know). There are no absolute laws, neither natural or divine, from which you can deduce universal criteria that unambiguously separate good from bad. It is always the individual that needs to make a specific choice in a specific situation.

There are however evolutionary *values* that can guide us in these decisions. Choices are thus not purely subjective or random. From the basic value of fitness, some more concrete, general values can be deduced:

- stability, equilibrium, robustness, sustainability: aim to make your choices durable by creating system that can withstand a variety of disturbances
- variation, innovation, exploration, experimentation: dare to go beyond what you already know, and try out new things
- diversity, variety: be aware that a system with more diversity is also more adaptive, robust and creative
- autonomy, self-organisation: do not try to direct or control specific individuals or systems; it is much more efficient to let them find their own way
- adaptation, integration in the larger whole: take into account that no individual or system can exist on its own; it always need to fit into its environment.
- efficiency, reduction of friction: take into account that energy and resources tend to dissipate, and therefore become unusable; design systems such that this waste is minimized

Specifically for goal-directed systems, and therefore also for human individuals and societies, we can note some additional values:

- insight in own goals or preferences: try to understand what is really important or valuable for you
- control, management of one’s own situation: develop the capability to counteract disturbances that drive you away from your goals

- reserves, buffers: make sure that you have enough resources to absorb unexpected fluctuations
- sensitivity, perception: develop your awareness of the different phenomena that may signal dangers or opportunities, even when these are still difficult to observe
- knowledge, intelligence: develop the capability to interpret the phenomena you sense, to infer their implications, and to devise appropriate plans to deal with them
- power, energy: make sure you have the physical capacity to act on the problems you sense

- What is true and what is false?

There are no *absolute* truths. The "truth" of a theory or model is nothing more than its ability to make predictions that are confirmed in practice. Two different theories can however make similar predictions without the one being true and the other false. These theories can be compared with two systems that have adapted in different ways to the same reality. Both are equally fit, but they are very differently organised.

Truth is however not only relative, subjective or culture-dependent. There is definitely a difference between a theory that makes reliable predictions (e.g. astronomy), and a theory that does not (e.g. astrology). Of two systems that compete for the same niche, the one will in general be more fit than the other, and there are certainly objective differences, advantages or disadvantages. These enable elimination ("falsification") of bad theories.

- Why do we die?

Biological fitness is reached through a combination of survival and reproduction. Every living system has a finite amount of energy to invest in fitness. What is used for the one (for example reproduction), is no longer available for the other (for example survival) (see 11.3). Therefore, there is a trade-off between survival (K-strategy) and reproduction (r-strategy). Since it is impossible to guarantee survival in every situation (because of for example accidents, illness, predators ...) sufficient energy has to be invested in reproduction to compensate for the inevitable loss of lives. This is why our organism will not invest maximally in survival in the long term. The result is aging and decline when the fertile period of reproduction is over. This evolutionary explanation for aging is the theory of the *disposable soma* (disposable body): it is most important that our genes live on in our descendants; our individual organism ("*soma*") is secondary and can be sacrificed.

- How can we be happy?

Happiness, meaning positive feelings or subjective wellbeing, is the biologically evolved signal that all is well, that the organism can continue like this. This means that the organism is fit, and able to cope with all practical foreseeable problems. In other words, that the organism has control over its situation and can attain its goals. It is not so much the external, objective situation that produces happiness, but the perception or the feeling that one has control over the situation, that one makes progress towards one's goals, and that there are no insurmountable problems. This explains why so many people who appear to have everything needed to be happy, still can be depressed or even commit suicide.

Over the long term, however, happiness requires a number of objective basic conditions: *health*, *knowledge*, social *participation* (personal relationships, being accepted within a group), *freedom* (making your own choices), *equality* (not being discriminated against), *prosperity*, and *safety* (low risk of accidents, crime, war, ...). That these factors contribute to happiness has been shown through a very extensive body of empirical research (as collected in Ruut Veenhoven's *World Database of Happiness*).

These conditions can however also be theoretically deduced from the general evolutionary-systemic values above, and the specific features of humans as social animals. Health indicates for example internal fitness, safety the absence of too strong perturbations, and prosperity the presence of all required resources. For their part, freedom and equality indicate the absence of suppression by dominant individuals or subgroups, and participation indicates cooperation with and support from others.

- What is the meaning of life?

"Mother, what are we living for?" The essence of the ESW can be summarised in one phrase: *the meaning of life is striving for fitness*. Fitness is the implicit goal of all systems. In practice, however, we still do not know what the best way is to increase fitness for a given system. Fitness itself is a very abstract, complex and multidimensional concept. There are a large variety of ways to increase fitness. Therefore, individuals still need to make their own choices, depending on their specific situation. But the ESW undoubtedly offers us a collection of useful guidelines.

Bibliography

Works for a general audience

The following publications are meant for a non-academic audience, but have mostly been written by scientists (with the exception of the journalists Gleick, Waldrop and Kelly), which guarantees a correct and thorough treatment of the subjects. The number of asterisks (*) indicates how recommendable or relevant the book is for readers of this book. Publications with a website <http://...> are available for free download at that address.

Aerts, D., Apostel L., De Moor B., Hellemans S., Maex E., Van Belle *., Van Der Veken J. (1994): Worldviews: From Fragmentation to Integration, (VUB Press, Brussels). http://pcp.vub.ac.be/books/Worldviews.html **

A short introduction to the worldview concept and its importance, with a series of suggestions for further research, among others about the ESW.

Casti J.L. Complexification: explaining a paradoxical world through the science of surprise, HarperCollins, 1994.

General introduction to a number of mostly mathematical concepts and observations that are important for the ESW, such as chaos, cellular automata and catastrophes.

Cziko G. Without Miracles: Universal Selection Theory and the Second Darwinian Revolution (MIT Press). http://faculty.ed.uiuc.edu/g-cziko/wm/ **

Variation and selection as universal explanatory model, applied to inter alia biological evolution, the immune system, knowledge, and control.

Davies, Paul: The Cosmic Blueprint: New Discoveries in Nature's Creative Ability to Order the Universe, 1988 Simon & Schuster. ***

An excellent and easy to understand overview of fundamental insights and phenomena around self-organisation, however with the ontological conclusion that there might be a "plan" after all, that sends evolution in the right direction.

Dawkins R. The Blind Watchmaker, Longman, London, 1986 ***

A well-written argumentation with plenty of biological examples about the force of the traditional, Darwinian theory of evolution as explanation for biological organisation.

Dawkins R. The selfish gene (2nd edition), Oxford University Press, Oxford, 1989 ***

A fascinating demonstration, albeit somewhat reductionist, about how the evolution of "replicators" (genes or memes) directs animal and human behaviour; possibly the most exciting scientific book available.

de Duve, Christian: Vital Dust: life as a cosmic imperative, BasicBooks, 1995. *

A Belgian winner of the Nobel Price Physiology delivers a detailed history of the origin and evolution of life, with a focus on the increase of complexity, albeit without a lot of reference to the ESW frame of reference (see also the more technical books of Maynard Smith and Szathmary).

de Rosnay, Joël: Le Macroscope / The Macroscope: a new scientific worldview, Harper & Row, New York, 1979. http://pespmc1.vub.ac.be/MACRBOOK.html ***

An easy-to-read, broad and thorough introduction into the ESW, with applications to the ecosystem and the economy of the world as a whole.

de Rosnay, Joël : L'Homme Symbiotique. Regards sur le troisième millénaire (Seuil, Paris, 1996) / The Symbiotic Man: A New Understanding of the Organization of Life and a Vision of the Future ** http://pespmc1.vub.ac.be/books/DeRosnay.TheSymbioticMan.pdf

Application of the ESW to better understand the present and future development of technology and society.

Gell-Mann, M., The Quark and the Jaguar: Adventures in the Simple and the Complex, W.*. Freeman, San Francisco, 1994.

A Nobel Price Physics winner and co-founder of the Santa Fe Institute describes his intellectual development and explains a couple of core concepts of the ESW.

Gleick, J. 1987. Chaos: Making of a New Science, Penguin Books, New York.

Biography about how a number of scientists developed chaos theory, unfortunately with little explanation about the actual ideas.

Heylighen F., Joslyn C. & Turchin V. (eds.) (1993-2004): "Principia Cybernetica Web" (http://pcp.vub.ac.be/). ***

By far the most important website about the ESW, with a wealth of information, but a rather complex structure.

Kauffman S. A. At Home in the Universe: The Search for Laws of Self-Organization and Complexity, Oxford University Press, Oxford, 1995. *

Popularised version of Kauffman's most important insights and CAS simulations around the origin and evolution of life, but a bit disappointing in style and organisation.

Kelly, Kevin: Out of Control: The New Biology of Machines, Social Systems and the Economic World (Perseus Press, 1995) http://www.kk.org/outofcontrol/contents.html ***

A very nice introduction to the ideas, history and future development of the ESW, full of interesting anecdotes, and with emphasis on applications in information technology and network economy.

Laszlo, Ervin: The Systems View of the World: The Natural Philosophy of the New Developments in the Sciences (1995 George Braziller) *

A short, philosophical introduction to system approach.

Maturana H. R., & Varela F. J. The Tree of Knowledge: The Biological Roots of Understanding, (rev. ed.), Shambhala, Boston, 1992.

Popularised version of the important but complicated theory of autopoiesis and its implications for epistemology.

Prigogine, I. and Stengers, I. La Nouvelle Alliance / Order Out of Chaos, Bantam Books, New York, 1984 **

Scientific and philosophical implications of Prigogine's insights about self-organisation, and especially about how these undermine the Newtonian worldview.

Simon, *. A. The Sciences of the Artificial (3rd. edition) MIT Press, Cambridge MA, 1996.**

Problem solving through heuristic trial-and-error as a universal mechanism to adapt to a complex environment, with applications for psychology, computing, economy, organisations and the evolution of complexity.

Stewart, John E. Evolution's Arrow: The direction of evolution and the future of humanity (Chapman Press, Australia, 2000): http://www4.tpg.com.au/users/jes999/ **

A thorough argumentation that evolution leads to cooperative systems on ever-greater scale and with ever more adaptivity.

Turchin, Valentin: The Phenomenon of Science. A cybernetic approach to human evolution 1977 (Columbia University Press, New York). http://pcp.vub.ac.be/POSBOOK.html **

Introduction to and of the theory of metasystems transitions, with applications mostly in the historical development of scientific knowledge.

Waldrop M. M. Complexity: The Emerging Science at the Edge of Order and Chaos, Simon & Schuster, New York, 1992 **

How the scientists that laid the foundations of the CAS approach and the Santa Fe Institute came to their most important insights.

Wright, Robert. Non-Zero. The Logic of Human Destiny (Pantheon Books, 2000) **

A very well written book that applies certain core concepts of the ESW (comparable with those of Stewart) to explain the (pre)historical development of civilisation (and to a lesser degree that of biological organisation).

Technical works

The publications below are aimed at a more restricted audience, but do in general not require too much specialised knowledge.

Anderson P. W., K. J. Arrow, and D. Pines (Eds.). The Economy as an Evolving Complex System, Addison-Wesley, Redwood City CA, 1988.

Arthur, W. B.: Positive Feedbacks in the Economy, Scientific American, February 1990, pp. 92-99.

Ashby W. R. An Introduction to Cybernetics, Methuen, London, 1964. *** http://pcp.vub.ac.be/ASHBBOOK.html

Ashby, W. R. Design for a Brain - The Origin of Adaptive Behaviour. Chapman and Hall, London, 1960.

Axelrod R. M. The Evolution of Cooperation, Basic Books, New York, 1984. **

Ball, P. (2012). Why Society is a Complex Matter: Meeting Twenty-first Century Challenges with a New Kind of Science (2012 edition.). New York: Springer.

Bak P. and Chen K.: Self-Organized Criticality, Scientific American: January 1991, pp. 46-53.

Bak P., How Nature Works: The Science of Self-Organized Criticality, Springer, Berlin, 1996.

Bateson, Gregory: (1972) Steps to an Ecology of Mind, Ballantine, New York **

Bennett C. *. Dissipation, Information, Computational Complexity and the Definition of Organization. Emerging Syntheses in Science, Pines D. (ed.), Addison-Wesley, Redwood City CA, 1985, pp. 215-233. *

Boulding, Ken: (1956) "General Systems Theory - The Skeleton of Science", General Systems Yearbook, v. 1, pp. 11-17 **

Campbell, D. T. "Downward Causation" in Hierarchically Organized Biological Systems. Studies in the Philosophy of Biology, F.J. Ayala and T. Dobzhansky (eds.), Macmillan, New York, 1974 . *

Campbell, D. T. Evolutionary epistemology. in: Evolutionary epistemology, rationality, and the sociology of knowledge, G. Radnitzky and W. W. Bartley (eds.), Open Court, La Salle IL, 1987, pp. 47-89. **

Checkland, Peter: (1981) Systems Thinking, Systems Practice, Wiley, New York *

Crutchfield, J., Farmer, J.D., Packard, N., and Shaw, R.: Chaos, Scientific American, 255 (6): December 1986, pp. 46-57.

Darwin C. The origin of species by means of natural selection or the preservation of favoured races in the struggle for life. (Edited with and introduction by J. W. Burrow). Penguin classics, 1985. (First published by John Murray, 1859) **

Dennett D. (1996) Darwin's dangerous idea: Evolution and the Meanings of Life (new edition, Penguin Books). **

Eigen M. and P. Schuster. The Hypercycle: A principle of natural self-organization, Springer, Berlin, 1979

Eigen M., and R. Winkler-Oswatitsch. Steps Toward Life: a perspective on evolution. Oxford University Press, New York, 1992.

Flood, Robert, and Ewart Carson. (1993). Dealing with Complexity: An Introduction to the Theory and Application of Systems Science. New York: Plenum. *

Gershenson, C., & Heylighen, F. (2005). How can we think complex? In K. Richardson (Ed.), Managing organizational complexity: philosophy, theory and application (pp. 47–61). Information Age Publishing. Retrieved from http://pcp.vub.ac.be/Papers/ThinkingComplex.pdf

Haken *. Synergetics: an introduction, Springer, Berlin, 1978. **

Heylighen F. & Joslyn C. (2001): “Cybernetics and Second Order Cybernetics”, in: R.A. Meyers (ed.), *Encyclopedia of Physical Science & Technology* (3rd ed.), Vol. 4 , (Academic Press, New York), p. 155-170. **

Heylighen F. (1997): “Publications on Complex, Evolving Systems: a citation-based survey”, *Complexity* 2 (5), p. 31-36. *

Heylighen F. (1999): “The Science of Self-organization and Adaptivity”, in: *The Encyclopedia of Life Support Systems* (EOLSS Publishers Co. Ltd) [in press]**

Heylighen F. (2000): “Evolutionary Transitions: how do levels of complexity emerge?”, *Complexity* 6 (1), p. 53-57

Heylighen F. (2000): “Foundations and Methodology for an Evolutionary Worldview: a review of the Principia Cybernetica Project”, *Foundations of Science,* 5, p. 457-490. **

Heylighen, F. (2008). Complexity and Self-organization. *Encyclopedia of Library and Information Sciences, Eds. MJ Bates & MN Maack Taylor & Francis*. Retrieved from http://pespmc1.vub.ac.be/Papers/ELIS-complexity.pdf

Heylighen, F. (2011). Self-organization of complex, intelligent systems: an action ontology for transdisciplinary integration. *Integral Review*. Retrieved from http://pespmc1.vub.ac.be/Papers/ECCO-paradigm.pdf

Heylighen, F., Cilliers, P., & Gershenson, C. (2007). Complexity and Philosophy. In J. Bogg & R. Geyer (Eds.), *Complexity, science and society* (pp. 117–134). Oxford: Radcliffe Publishing,. Retrieved from http://arxiv.org/abs/cs/0604072

Heylighen, F., & Joslyn, C. (2003). Cybernetics and Second-Order Cybernetics. In Robert A. Meyers (Ed.), *Encyclopedia of Physical Science and Technology* (3rd ed., Vol. 4, pp. 155–169). New York: Academic Press. Retrieved from http://pespmc1.vub.ac.be/Papers/cybernetics-EPST.pdf

Heylighen F., Bollen J. & Riegler A. (eds.) (1999): *The Evolution of Complexity* (Kluwer Academic, Dordrecht). **

Holland, J. (1996). *Hidden Order: How Adaptation Builds Complexity*. Reading, Mass.: Basic Books.

Holland, J. H. (2012). *Signals and Boundaries: Building Blocks for Complex Adaptive Systems*. Cambridge, Mass: The MIT Press.

Holland, J. H. (2014). *Complexity: A Very Short Introduction*. OUP Oxford.
Jantsch, E., The Self-Organizing Universe: Scientific and Human Implications of the Emerging Paradigm of Evolution, Oxford, Pergamon Press, 1979.*

Kauffman S. A. The Origins of Order: Self-Organization and Selection in Evolution, Oxford University Press, New York, 1993 *

Kauffman S. A. Antichaos and Adaptation, Scientific American: August 1991, pp. 78-84 **

Klir, George: (1992) Facets of Systems Science, Plenum, New York **

Langton, C. G. (ed.), Artificial Life: An Overview, MIT Press, Cambridge, MA, 1995.

Laszlo, Ervin: Evolution, the Grand Synthesis 1987) Shambhala *

A holistic theory of evolution on all levels, physical, biological, mental.

Laszlo, E. (1996). *The Systems View of the World: A Holistic Vision for Our Time* (2nd edition.). Cresskill, NJ: Hampton Pr.

Maynard Smith, John and Eörs Szathmáry (1999): The Origins of Life: From the Birth of Life to the Origin of Language (Oxford University Press). **

Miller, J. H., & Page, S. E. (2007). *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton, N.J: Princeton University Press.

Mitchell, M. (2011). *Complexity: A Guided Tour*. Oxford; New York: Oxford University Press.
Monod, J. Le hasard et la nécessité / Chance and Necessity, Collins, London, 1972.

Morin, Edgar (1977) La méthode: I. La nature de la nature (Editions du Seuil).*

Morin, Edgar "*Introduction à la pensée complexe*". ESF Editeur, Coll. Communication et complexité, Paris, 1990.

Nicolis, G, and Prigogine, I. Self-Organization in Non-Equilibrium Systems, Wiley, New York, 1977.

Nicolis, G. and I. Prigogine. Exploring Complexity, Freeman, New York, 1989. *

Pettersson, Max (1996): Complexity and Evolution (Cambridge University Press)

Prigogine, I. and Stengers, I. Entre le temps et l'éternité / The End of Certainty: Time, Chaos, and the New Laws of Nature

Prigogine, I. From Being to Becoming: Time and complexity in the physical sciences, Freeman, San Francisco, 1980.

Shannon, C. E., and W. Weaver. The Mathematical Theory of Communication (5th ed.). University of Illinois Press, Chicago, 1963.

Veld, J. in 't (1998). Analyse van organisatieproblemen: Een toepassing van denken in systemen en processen. Leiden: Stenfert Kroese. *

von Bertalanffy L. General Systems Theory (Revised Edition), George Braziller, New York, 1973. ***

von Foerster H. Cybernetics of Cybernetics (2nd edition). Future Systems, Minneapolis, 1996. **

von Foerster H. Observing Systems: Selected papers of Heinz von Foerster. Intersystems, Seaside, CA, 1981. *

von Foerster H. On self-organising systems and their environments. in: Self-Organising Systems, M.C. Yovits and S. Cameron (Eds.), Pergamon Press, London, 1960, pp. 30-50. **

Weinberg, G. M. (2001). *An Introduction to General Systems Thinking* (25 Anv edition.). New York: Dorset House.

Wiener N. Cybernetics: or Control and Communication in the Animal and Machine, M.I.T. Press, New York, 1961. *

Wolfram S. Cellular Automata and Complexity: Collected Papers, Addison-Wesley, Reading MA, 1994.

Index